

Tartu Ülikool
Arvutiteaduse instituut
Infotehnoloogia mitteinformaatikutele õppekava

Teet Kalmus

**Metoodika kasvatusväärtuste automaatseks
eraldamiseks *online*-meediatekstidest**

Magistritöö (15 EAP)

Juhendajad: Dage Särg, MA
prof. Veronika Kalmus

Tartu 2019

Metoodika kasvatusväärtuste automaatseks eraldamiseks *online*-meediatekstidest

Lühikokkuvõte:

Magistritöö eesmärgiks oli luua metoodika koos tarkvaralahendusega, mis võimaldaks *online*-meediatekstide hulgast leida üles tekstid, mis sisaldavad kasvatusväärtusi, ning neid tekste analüüsida.

Loodud tarkvaralahendus hõlmab andmete leidmist ja allalaadimist Perekooli Koolilapse foorumi meediatekstidest, kasvatusväärtuste sõnastiku loomist, kasvatusväärtusi sisaldavate tekstide filtreerimist ning analüüsi, sh ka tekstide emotsionaalsuse analüüsi.

Tarkvaralahenduse loomisel on kasutatud Tartu ülikoolis loodud keeletöötlusteeki EstNLTK ning magistritöö analüüsi käigus saadud andmeid on plaanis kasutada teadusartiklite kirjutamisel. Loodud metoodika alusel on plaanis analüüsida ka teisi meediatekste.

Võtmesõnad: emotsioonianalüüs, kasvatusväärtused, keeletöötlus, veebiämblik

CERCS: P175 Informaatika, süsteemiteooria

Methodology for Automated Extraction of Socialization Values from Online Media Texts

Abstract:

The aim of the present thesis was to create a methodology with a software solution to find texts containing socialization values among online media texts and to analyse the filtered texts.

The software solution includes web crawling and downloading the texts from the forum Koolilaps of Perekool, creating a dictionary of socialization values, filtering texts using the dictionary of socialization values and analysing the filtered texts with sentiment analysis.

The software package EstNLTK was used to create the software solution. Further plans include using the results of the thesis for writing academic articles and using the methodology and the software solution for analysing other media texts.

Keywords: sentiment analysis, socialization values, natural language processing, web crawling

CERCS: P175 Informatics, systems theory

Sisukord

1. Sissejuhatus	5
2. Teoreetilised lähtekohad.....	6
2.1 Väärtused ja kasvatusväärtused.....	6
2.1.1 Väärtused	6
2.1.2 Kasvatusväärtused.....	7
2.2 Emotsioonianalüüs	9
2.2.1 Emotsioonianalüüsi mõiste ja eesmärgid.....	9
2.2.2 Emotsioonianalüüsi meetodid.....	10
2.2.3 Emotsioonianalüüs eesti keeles	13
3. Tekstide filtreerimise metoodika	15
3.1 Metoodika kirjeldus.....	15
3.2 Uurimuse objekt ja hüpoteesid	16
3.2.1 Uurimuse objekt.....	16
3.2.2 Hüpoteesid	16
3.3 Teek EstNLTK	17
3.4 Kasvatusväärtuste sõnastiku loomine	19
3.4.1 Kasvatusväärtuste baassõnastiku loomine	19
3.4.2 Kasvatusväärtuste sünonüümide otsimine	20
3.5 Veebiämbliku tarkvara loomine ja andmete allalaadimine Perekooli foorumist.....	21
3.5.1 Tarkvara algoritmi koostamine	21
3.5.2 Tarkvara loomine ja testimine.....	22
3.5.3 Andmete allalaadimine	23
3.6 Kasvatusväärtusi sisaldavate tekstide filtreerimine	24
4. Analüüs ja tulemused	28
4.1 Emotsiooniskooride arvutamine	28
4.2 Tekstide arv, emotsiooniskoorid ja kirjutamise aeg.....	30
4.3 Tekstide pikkus ja keskmine emotsiooniskoor	32
4.4 Kasvatusväärtuste arv	33
4.5 Kasvatusväärtuste väärtusgruppide ja väärtusjaotuse keskmine emotsiooniskoor	35
4.6 Võimalus tekste kvalitatiivselt analüüsida	39
5. Kokkuvõte.....	41

6. Viidatud kirjanduse loetelu	42
Lisad.....	44
Lisa 1. Programmeerimiseks vajaliku keskkonna seadistamine	44
Lisa 2. Veebiämbliku mõiste ja kasutamise viisakusreeglid	46
Lisa 3. Küsimustik emadele, lehekülg 3 (Tulviste ja Mizera, 2010)	47
Lisa 4. Kasvatusväärtuste sõnastik	48
Litsents	49

1. Sissejuhatus

Milton Rokeach on defineerinud väärtusi kui kehtvat uskumust, et teatud liiki käitumine või olemise viis on personaalselt või sotsiaalselt eelistatum kui vastupidine või teistsugune käitumise liik või olemise viis (Rokeach, 1973). Shalom H. Schwartzi ja Wolfgang Bilsky määratluse kohaselt on väärtused üldistused või uskumused, mis on seotud soovitava käitumise või lõpptulemusega, mis on konkreetsete situatsioonide ülesed, suunavad käitumise ja sündmuste valikut ning on järjestatud suhtelise tähtsuse järgi (Schwartz ja Bilsky, 1987). Kasvatuseväärtused on väärtuste alamliik, kus fookus on just laste kasvatamisel kasutatavatel väärtustel.

Kasvatuseväärtuste kui olulise kultuuriindikaatori uurimine on siiani põhinenud peamiselt sotsiaalteaduslikel meetodidel, kus teadustööks vajalikke andmeid on kogutud peamiselt intervjuusid ja küsitlusi tehes. Kvantitatiivsete tekstianalüüsi meetodite poolelt on väärtuste analüüsil kasutatud sisuanalüüsi (kontentanalüüsi), mis on aga ajamahuks meetod, sest väärtusi sisaldavad tekstid tuleb üksikhaaval üles otsida ja käsitsi kodeerida.

Elame ajaperioodil, kus meediasisu loomine ei ole enam meediaettevõtete monopol, vaid läbi erinevate sotsiaalvõrgustike ja teiste veebiplatvormide saavad meediatekste luua kõik inimesed. See on kaasa toonud ühelt poolt arvamuste paljususe, aga teiselt poolt on sotsiaalmeedias ja virtuaalses kultuuriruumis loodud meediasisu läinud nii mahukaks, et sisuanalüüsiga on raske nii suurt hulka tekste analüüsida.

Lisaks kasvatuseväärtuste esinemissageduse määramisele on oluline vaadelda ka seda, missuguses emotsionaalses kontekstis neid esitatakse. Seda võimaldab tekstide emotsioonianalüüs.

Magistritöö eesmärgiks on Perekooli veebilehe foorumi „Laps“ alamfoorumi „Koolilaps“¹ baasil luua metoodika ja tarkvaralahendus, mida kasutades on võimalik salvestada andmed veebilehelt faili ja automaatselt leida salvestatud andmetest üles kasvatuseväärtused, arvutada tekstide emotsiooniskoorid ning siis neid andmeid analüüsida ja luua võimalused tekstide täiendavaks kvalitatiivseks analüüsiks.

Magistritöö esimene peatükk käsitleb kasvatuseväärtusi ja tekstide emotsionaalsuse analüüsi läbi teooriaprisma, teine peatükk käsitleb metoodikat kasvatuseväärtuste leidmiseks ja kolmas peatükk käsitleb kasvatuseväärtusi sisaldavate tekstide analüüsi ja tulemusi. Lisades on juhised tarkvaralahenduse kasutamiseks vajaliku keskkonna seadistamiseks, kasvatuseväärtuste leksikon, täiendavad selgitused veebiämbliku kohta ja kasvatuseväärtusi sisaldav lehekülg Tiia Tulviste ja Luule Mizera poolt loodud ankeedist (2010).

Magistritöö autor tänab oma juhendajaid ja prof. Tiia Tulvistet suurepärase koostöö eest.

¹<https://foorum.perekool.ee/foorum/koolilaps/>

2. Teoreetilised lähtekohad

Magistritöös kasutatud kasvatusväärtuste klassifikatsioon põhineb suures osas Shalom H. Schwartzi teadusartiklis *An Overview of the Schwartz Theory of Basic Values* esitatud väärtusklastritel (Schwartz, 2012).

2.1 Väärtused ja kasvatusväärtused

2.1.1 Väärtused

Väärtuste defineerimisel tõi Schwartz (2012) välja kuus peamist omadust, mis iseloomustavad väärtusi:

1. Väärtused on uskumused, mis on lahutamatult seotud emotsioonidega. Näiteks inimesed, kelle jaoks on väärtusena tähtis „sõltumatus“, muutuvad rahutuks, kui sõltumatus satub ohtu, nad muutuvad meeleheitel olevaks, kui ei suuda seda väärtust kaitsta ja on õnnelikud, kui saavad seda nautida.
2. Väärtused on motivaatoriteks eesmärkide saavutamisel. Näiteks inimesed, kelle jaoks on väärtustena olulised „sotsiaalne kord“, „õiglus“ ja „abivalmidus“, on motiveeritud neid väärtusi järgima oma eesmärkide saavutamisel.
3. Väärtused on konkreetsete olukordade ja tegude ülesed, nad on abstraktsed eesmärgid. Näiteks „sõnakuulekus“ ja „ausus“ väärtustena võivad olla olulised nii tööl, koolis, kodus kui sõprade või võõrastega koos olles. See eristab väärtusi normidest ja hoiakutest, mis on seotud konkreetsete tegude, objektide või olukordadega.
4. Väärtused juhivad inimesi otsuste tegemisel nagu standardid või kriteeriumid. Näiteks otsustavad inimesed väärtuste järgi, mis on halb või hea, mida sobib teha ja mida mitte. Tavaelus otsuste tegemisel inimesed üldjuhul väärtuste peale ei mõtle, väärtused tulevad meie teadvusesse siis, kui otsuste tegemisel lähevad väärtused omavahel konflikti ja tuleb teha valik nende vahel.
5. Väärtused on inimeste jaoks hierarhilises tähtsuse järjekorras. Just selline väärtussüsteem loob inimeste individuaalsuse. Näiteks, kas olulisemad väärtused on „saavutamine“ või „õiglus“, kas „uuendusmeelsus“ või „traditsioonid“? Väärtuste hierarhia eristab väärtusi hoiakutest ja normidest.
6. Inimesed juhinduvad oma käitumises väärtuste hierarhiast. Iga käitumine on üldjuhul mitme hierarhias erinevatel kõrgustel asuvate väärtuste koosmõju tulemus. Väärtused mõjutavad inimese tegutsemist, kui nad on aktiveeritud ja inimese jaoks antud sündmuse kontekstis olulised.

Erinevates riikides läbiviidud uuringute tulemusena koostas Schwartz kümnest klastrist koosneva indiviiditasandi väärtuste süsteemi, kus iga väärtusklastri alla kuuluvad veel üksikväärtused (Schwartz, 2012):

1. Iseseisvus – iseseisev mõtlemine, loomine, avastamine;
2. Stimulatsioon - põnevus, uudsus, väljakutsed elus;
3. Hedonism – füüsiline või vaimne rahuldus või rahulolu üksikisiku tasandil;
4. Saavutamine – edu personaalsel tasandil läbi kompetentsuse demonstreerimise vastavalt sotsiaalsetele standarditele;
5. Võim – sotsiaalne staatus ja prestiiž, kontroll või domineerimine inimeste ja ressursside üle;
6. Turvalisus – ohutus, harmoonia ja stabiilsus ühiskonna, suhete ja iseenda jaoks;
7. Konformism – tegude, tahte ja impulsside vaoshoidmine vältimaks teiste ärritamist või neile haiget tegemist ja vältimaks sotsiaalsete ootuste või normide rikkumist;
8. Traditsioonid – kultuuri või religiooni tavade ja normide austamine, heakskiitmine ja vastavalt nendele käitumine;
9. Heatahtlikkus - Heade suhete säilitamine ja parandamine üksikisiku tasandil nendega, kellega ollakse isiklikult kontaktis;
10. Universalism – inimeste ja looduse mõistmine, tolereerimine ja kaitsmine.

Schwartzi väärtuste süsteemist on lähtunud ka paljud kasvatusväärtuste uurijad.

2.1.2 Kasvatusväärtused

Võttes aluseks Rokeachi (1973) ning Smithi ja Schwartzi (1997) väärtuste definitsioonid, defineeris Veronika Kalmus kasvatusväärtusi kui kultuuri elementi, sotsiaalselt jagatud ettekujutust soovitatavatest omadustest, mida ühiskond püüab noores põlvkonnas sotsialiseerimise käigus kujundada (Kalmus, 2019).

Kasvatusväärtusi on jaotatud peamiselt kahte suurde gruppi, samas on neid erinevalt nimetatud.

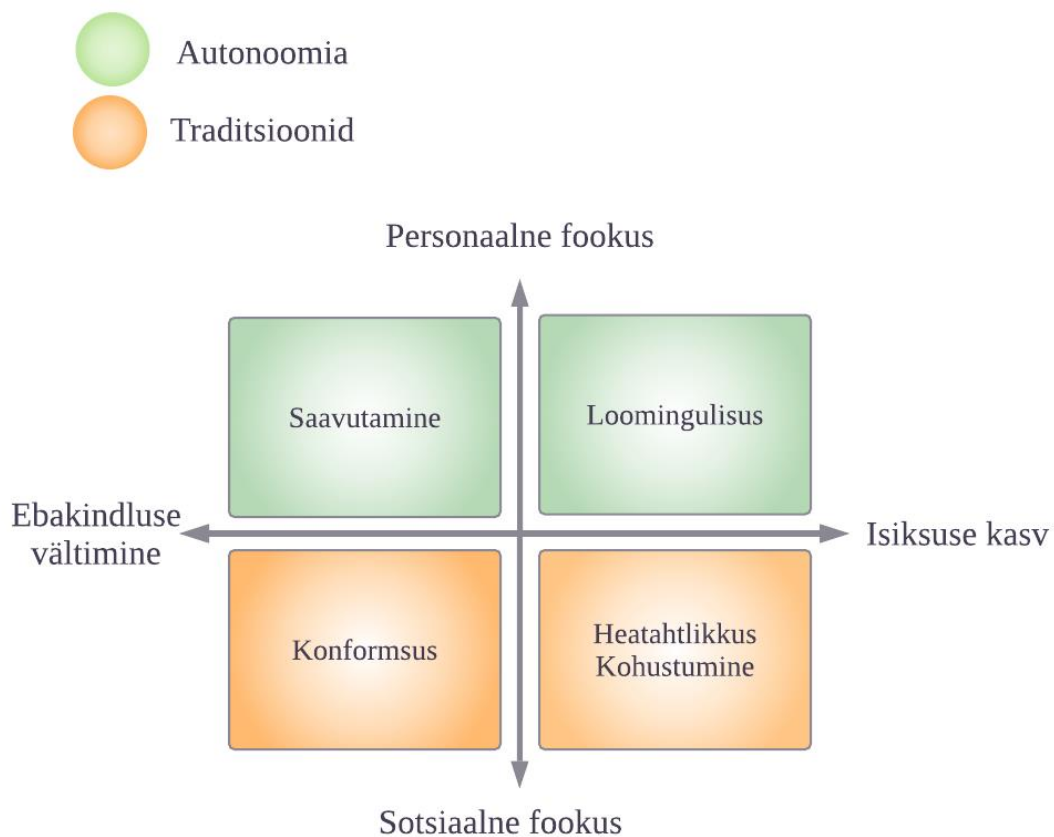
Melvin L. Kohn keskendus 1960ndatel erinevate sotsiaalsete klasside poolt eelistatud kasvatusväärtuste uurimisele ja ta jagas kasvatusväärtused kahte suuremasse rühma – iseseisev mõtlemine-otsustamine (*self-direction*) ja autoriteedile suunatud konformsus (*conformity to external authority*). Uuringute tulemusena leidis ta, et tööliklass eelistab pigem konformseid ja keskklass iseseisvale mõtlemisele-otsustamisele suunatud kasvatusväärtusi (Kohn, 1969).

Duane F. Alwin jagas kasvatusväärtused autonoomseteks (*autonomy*) ja alluvuslikeks (*obedience*) ehk siis „iseseisev mõtlemine“ versus „vanemate autoriteedile allumine“ (Alwin, 2001).

Tormod Fjellvang jagas kasvatusväärtused kahe dimensiooni vahel – autonoomia dimensioon ja traditsionalismi dimensioon (Fjellvang, 2011).

Käesolevas magistritöös on kasvatusväärtuse klassifitseerimisel aluseks võetud Schwartzi dünaamilise struktuuriga väärtuste jaotus (Schwartz, 2012), mille modifitseerimise järel on

meil kaks suurt kasvatusväärtuse klasterit (traditsiooniline ja autonoomne), mis mõlemad jagunevad kaheks alamklastriks (joonis 1).



Joonis 1. Modifitseeritud Schwartzi dünaamilise struktuuriga väärtuste jaotus (Schwartz, 2016)

Traditsiooniline klaster jaguneb „Konformsuseks“ (näiteks „kuulekas“, „viisakas“, „enese-distsipliin“) ja teiseks alamklastriks, mille moodustavad „Heatahtlikkus“ (näiteks „aus“, „abivalmis“, „andestav“) ja Schwartzi kultuuritasandi väärtusklastriks (Schwartz, 1994) üle võetud „Võrdväärne kohustumine“ (eestikeelne mõiste Kalmus, 1999), mida käesolevas magistritöös kasutakse väärtusklastriks „Kohustumine“ (näiteks „lojaalne“, „vastutus-tundlik“). Autonoomse klasteri moodustavad alamklastrid „Saavutamine“ (näiteks „ambit-sioonikas“, „ettevõtlik“, „edasipüüdlik“) ja „Loomingulisus“ (näiteks „loovus“, „iseseisvus“, „enesekindlus“).

Vertikaalteljel on dünaamika „Personaalne fookus“ *versus* „Sotsiaalne fookus“. Traditsioonilise väärtusklastri fookus on sotsiaalsusel ehk siis väärtused on suunatud grupi (ka ühiskonna) hüvanguks tegutsemisele (näiteks „abivalmis“) või siis ühiskondlikele normidele ja hoiakutele allumisele (näiteks „kuulekas“, „viisakas“). Autonoomse

väärtusklassi fookus on personaalsusel ehk siis on suunatud üksikisiku väärtustele, näiteks „ambitsioonikus“, „saavutusvajadus“, „otsustamisjulgus“.

Horisontaalteljel on dünaamika vasakult paremale „Ebakindluse vältimine“ *versus* „Isiksuse kasv“. Joonisel vasakul pool asuvate traditsioonilise väärtusklastri alamklastri „Konformsus“ ja autonoomse väärtusklastri alamklastri „Saavutamine“ väärtused on suunatud ebakindluse vältimisele kas aktiivselt ise tegutsedes sellel suunal (näiteks „ettevõtlikkus“) või siis maandades sisemist ebakindlust läbi konformsuse (näiteks „kuulekus“).

Joonisel paremal pool asuvate traditsioonilise väärtusklastri alamklastri „Heatahtlikkus/Kohustumine“ ja autonoomse väärtusklastri alamklastri „Loomingulisus“ väärtused on suunatud eeskätt isiksuse arendamisele läbi erinevate väärtuste nagu näiteks „loovus“ ja „andestamine“.

Schwartzi väärtusteooria järgi (Schwartz, 2012) on väärtused hierarhilises tähtsuse järjekorras ja pidevalt toimuvad erinevate väärtuste omavahelised interaktsioonid ning mida vastandlikumad on väärtused, seda suurem on väärtuskonflikt. Joonisel saab väärtused samamoodi jaotada hierarhiliselt kolmeks. Kõige sarnasemad on omavahel horisontaalteljel asuvad alamklastrid (joonisel 1 sama värvi kastid) ja inimene oma otsustes tihtipeale ei tajugi erinevust nendes väärtustes. Väärtuskonflikt on suurem vertikaalteljel kõrvuti asetsevate alamklastrite puhul („Saavutamine“ *versus* „Konformsus“ ja „Loomingulisus“ *versus* „Heatahtlikkus/kohustumine“), aga neid seovad siiski horisontaaltelje dünaamika mõisted, näiteks suure saavutusvajadusega inimene mõistab, et karjääri tegemiseks tuleb osata organisatsioonis mingi aja konformselt käituda. Või et isiksuse arendamisel ei ole halb ka mingil perioodil tegutseda suurema sotsiaalse fookusega väärtuste järgi.

Kõige suurem väärtuskonflikt on diagonaalidel asuvate alamklastrite vahel („Saavutamine“ *versus* „Heatahtlikkus/kohustumine“ ja „Loomingulisus“ *versus* „Konformsus“), näiteks on väga raske ühitada selliseid väärtusi nagu „kuulekus“ ja „loovus“ või „ettevõtlikkus“ ja „isetus“.

2.2 Emotsioonianalüüs

2.2.1 Emotsioonianalüüsi mõiste ja eesmärgid

Sotsiaalsõrgustike ja erinevate veebikeskkondade foorumite areng on loonud meediamaaistikul uudse olukorra, kus tekste luuakse suurtes kogustes ja erinevatel põhjustel on vaja neid tekste automatiseeritult analüüsida (näiteks ärilised ja teaduslikud), et leida sealt kasulikku informatsiooni.

Üheks levinumaks selliste tekstide automaatse analüüsimise tehnikaks on emotsioonianalüüs (eesti keeles veel ka „tundeanalüüs“, „meelestatuse analüüs“, inglise keeles „sentiment analysis“ või „opinion mining“), mida võib defineerida kui tegevust, mille

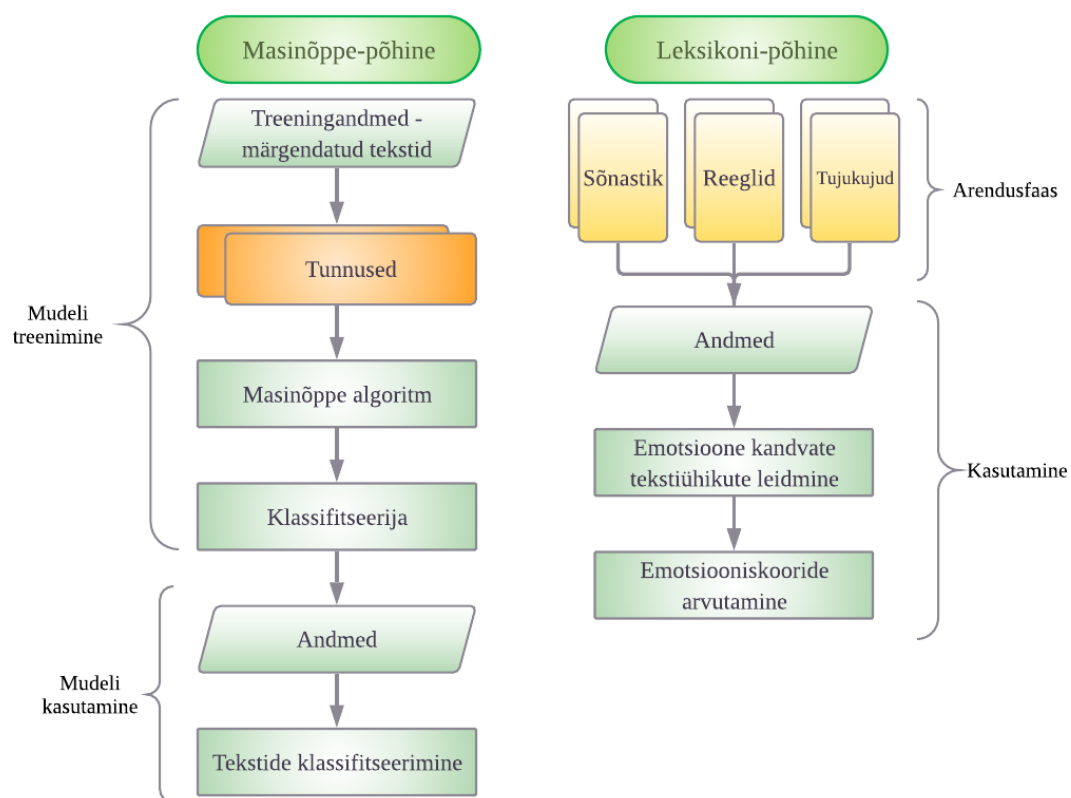
käigus automaatselt leitakse tekstidest emotsioone edasiandvad sõnad, sõnakombinatsioonid ja tujukujud (ehk emotikonid) ning saadud andmete põhjal tehakse otsus tekstide emotsionaalsuse osas.

Emotsioonianalüüsil on üldjuhul kolm eesmärki (Esuli ja Sebastiani, 2005):

- Teksti polaarsuse määramine skaalal subjektiivne-objektiivne;
- Teksti polaarsuse määramine skaalal positiivne-negatiivne;
- Teksti polaarsuse tugevuse määramine skaalal positiivne-negatiivne, näiteks kas tekstis väljendatud konkreetne emotsioon on nõrgalt positiivne, keskmiselt positiivne või tugevalt positiivne.

2.2.2 Emotsioonianalüüsi meetodid

Emotsioonianalüüsis on kaks peamist meetodit – leksikoni-põhine ja masinõppe-põhine (joonis 2).



Joonis 2. Masinõppe-põhine ja leksikoni-põhine emotsioonianalüüs (Taboada, 2016).

Leksikoni-põhises meetodis lähtutakse eeldusest, et tekst sisaldab sõnu, millel on kas negatiivne või positiivne emotsioon, näiteks sõna „hea“ on positiivse ja sõna „halb“ on negatiivse emotsiooniga. Leksikoni-põhine meetod eeldab sõnastiku olemasolu, kus on nii positiivse kui negatiivse emotsiooniga sõnad koos polaarsuse indikatsiooniga (näiteks on

positiivse sõna arvuline väärtus 1 ja negatiivsel -1). Leksikon võib sisaldada ka emotsiooni tugevust, selleks tuleb emotsioonid paigutada skaalale emotsiooni tugevuse järgi (Pajupuu, Altrov ja Pajupuu, 2016). Analüüsitavatest tekstidest otsitakse sõnu, mis on olemas emotsioonisõnastikus, ja sõnade leidmisel liidetakse emotsioonisõnastikus sõnadel olevad arvulised väärtused ning emotsiooniskoori põhjal võib teha järelduse teksti emotsionaalse polaarsuse kohta.

Masinõppe puhul kasutatakse enamasti juhendatud-õppe meetodit (*supervised learning*), st teksti polaarsuse määramisel kasutatakse klassifikaatoreid, mida on treenitud märgendatud tekstiühikutega, kus dokumendid või laused on määratletud kas positiivseteks ja negatiivseteks või on kolmene jaotus – positiivsed, negatiivsed ja neutraalsed (Taboada, 2016).

Kuna masinõppe meetod põhineb tekstimustrite tundmaõppimisel, siis on võimalikult hea tulemuse saavutamiseks oluline teksti puhastamine (näiteks tuleks eemaldada veebiaadressid, *hashtag*'id jne), samas leksikoni-põhises lähenemises see nii oluline ei ole, sest tekstimustrite asemel otsitakse emotsioone kandvaid sõnu, sõnakombinatsioone või tujukujusid (Taboada, 2016).

Mõlemad meetodid on andnud häid tulemusi ja tulemuste täpsus on ingliskeelsetes tekstides üldjuhul vahemikus 70-90%, kusjuures parimad tulemused on saavutatud kindla valdkonna tekstidega, näiteks filmide arvustuste analüüs koos binaarse klassifitseerimisega (Pang ja Lee, 2008).

Siiski on mõlema meetodi kasutamisel olemas probleemid ja nende probleemide lahendamisel ja leevendamisel kasutatakse erinevaid võtteid. Peamised probleemid seoses leksikoni põhise meetodi kasutamisega:

1. Leksikoni koostamine. Võiks arvata, et kui panna leksikoni võimalikult palju emotsiooni kandvaid sõnu, siis tekstide emotsionaalsuse määramise täpsus kasvab, aga uuringud näitavad, et mingist piirist suurema sõnade arvuga emotsionaalsussõnastik muudab tulemuse hoopis halvemaks, sest tekib müra, ning parema tulemuse saab väiksema sõnastikuga (Taboada jt, 2011). Põhjuseks asjaolu, et kuigi inimese sõnavaras on oma emakeeles rääkides üle 50000 sõna, kasutab ta igapäevaelus keskmiselt nendest ainult kümnendikku (Kwatler, 2017) ja kõige levinumate emotsiooni kandvate sõnade leksikoni panemine tagab hea tulemuse. Nii on laialdaselt kasutatavates ingliskeelsetes emotsioonisõnastikes suurusjärgus 5000-8000 sõna, sealt edasi sõnade arvu kasvatamine enam tulemust olulisel määral ei paranda (Taboada, 2016). Leksikoni koostamine nullist on küll suur töö, aga kõigil soovijatel on võimalik kasutada juba teadlaste poolt loodud emotsioonisõnastikke. Sõnad on ingliskeelsetes emotsioonisõnastikes üldjuhul lemmatiseeritult (algvormina) ja sõnaliikidest eelistatakse kasutada omadussõnu (Taboada, 2016).

2. Keele ülesehituse eripäradest tulenevalt kipuvad leksikoni kasutades emotsiooniskoorid tulema liialt positiivsed, sest keeles on otsese negatiivse emotsionaalsusega sõnu vähem kui positiivse emotsionaalsusega sõnu (Taboada, 2016). Mis ei tähenda, et negatiivse emotsionaalsusega tekste oleks vähem, aga nende täpsemaks määratlemiseks ei piisa ainult emotsioonileksikoni kasutamisest, vaid tuleb luua täiendavad reeglid.
3. Tekstide emotsionaalsusega seoses kasutatakse mõisteid teksti emotsionaalse tähenduse muutja (*valence shifter*, näiteks eituse) ning emotsionaalse tähenduse võimendajad (*intensifiers*) – suurendajad (*amplifiers*) ja vähendajad (*downtoners*) (Taboada jt, 2011). Võrdleme näiteks lauseid „Tõnu oli väga kuri Teedu peale“ ja „Tõnu oli natuke kuri Teedu peale“, kus „väga“ on suurendaja ja „natuke“ on vähendaja. Ainult emotsioonisõnade leksikoni kasutades on selliseid tähenduste muutjaid raske arvesse võtta.
4. Emotsionaalselt laetud tekste saab kirjutada nii, et seal üldse puuduvad emotsioonisõnastikus olevad sõnad või on nad tähenduselt täpselt vastupidise tähendusega. Nii ongi just leksikoni-põhise meetodi suurimaks puuduseks, et sellega on problemaatiline määratleda teksti emotsionaalsust juhtudel, kui kasutatakse irooniat ja sarkasmi ning seal pole abi ka täiendavate reeglite kasutamisest, sest erinevalt masinõppe meetodist ei saa leksikoni-põhises meetodis õpetada programmi tuvastama tekstide mustreid.
5. Veebitekstides kasutatakse tihtipeale emotsioonide edasiandmiseks väljendeid, mis on slängina või lühenditena, lisaks ebatsensuursed sõnad, mille roll teksti emotsionaalsuse muutmisel on tihtipeale ambivalentne.
6. Erinevate valdkondade tekstid on sisult niivõrd erinevad, et parema tulemuse saamiseks tuleb emotsioonisõnastikku teha muudatusi (Taboada, 2016). Samas, kuigi muudatuste tegemine emotsioonisõnastikku on märkimisväärne töö, on see lihtsam kui masinõppe-meetodil mudelit treenida täiendavate klassifitseerijatega (Brooke, Tofiloski ja Taboada, 2009).
7. Olemasolevad emotsioonileksikonid on peamiselt inglise keeles, nagu ka tarkvara, mis võimaldab teha nende baasilt kvaliteetsemat analüüsi, näiteks programmeerimiskeeles R pakett `qdap` (Kwatler, 2017). Teistes keeltes tähendab omakeelse emotsioonisõnastiku kasutamine seda, et tuleb osata küllaltki hästi programmeerida, saavutamaks paremat analüüsitulemust (näiteks reeglite lisamine).

Eelpoolnimetatud probleemide vähendamiseks on kasutusel mitmed lahendused:

- Reeglid ehk siis lisaks leksikoni kasutamisele täiendavate tegevuste sooritamine, mis muudab tulemust paremaks. Levinud meetod on lauses eituse esinemisel enne emotsionaalselt neutraalset tegusõna (mis ei sisaldu emotsioonisõnastikus) lisada see emotsionaalsusskoorile kui negatiivse emotsiooniga sõna, näiteks lause „Tõnu ei läinud täna kooli“ on negatiivse emotsionaalse laenguga, samas kui lause kooliminemise kohta oleks neutraalne. Veelgi suurem muutus toimub lausetes, kus

tegusõnad on selgelt emotsionaalsust väljendavad (Pajupuu jt, 2016). Näiteks lause - „Tõnule ei meeldi kevad“ - oleks ilma täiendavate reegliteta positiivse emotsiooniskooriga olukorras, kui tegusõna „meeldima“ sisalduks emotsioonisõnastikus. Sellisel juhul ei piisa ühe ühiku võrra skoori muutmisest, vaid reegel liidaks kaks vastasmärgiga ühikut emotsiooniskoorile, näiteks „ei meeldi“ puhul oleks see -2.

- Leksikoni täiendamine slängi- ja vandesõnadega. See tegevus vähendab sõnastiku universaalsust, sest släng ja vandesõnad erinevad oma tähenduselt erinevates tekstides, samas annab parema tulemuse just mingeid konkreetseid tekste analüüsid, näiteks ka Twitteri säutse või foorumite postitusi analüüsid.
- Leksikoni muutmine vastavalt analüüsitavate tekstide spetsiifikale ehk siis näiteks kasvatusväärtusi sisaldavates tekstides on palju kodu ja perekonnaga seotud sõnu, mis võivad olla emotsioonisõnastikus kui positiivse emotsiooniga sõnad (näiteks „ema“, „lapsepõlv“), aga antud kontekstis on nad neutraalsed ja positiivsena arvestamine tekitab valepositiivsust.

Masinõppepõhise-meetodiga tekstide emotsionaalsuse määramisel on oluline, kui hästi on mudel treenitud klassifitseerimaks tekste. Masinõpe põhineb tekstide mustrite tundmaõppimisel – mida rohkemate positiivse ja negatiivse emotsionaalsusega tekstidega mudel treenitakse, seda parem tulemus konkreetse valdkonna tekstide emotsionaalsuse määramisel, aga vaid üht liiki tekstidel treenitud mudel ei ole universaalne, vaid toimib hästi ainult sama liiki tekstidel – näiteks filmide kohta kirjutatud arvamuste emotsionaalsuse määramisel saavutatakse isegi täpsus 90% ja enamgi, aga sama mudeliga teiste valdkondade tekstide analüüsimisel on tulemus keskpärasem (Taboada, 2016).

2.2.3 Emotsioonianalüüs eesti keeles

Kuna emotsionaalsuse analüüs on väga suures osas ingliskeelne, siis otsustasid Eesti teadlased Hille Pajupuu, Rene Altrov ja Jaan Pajupuu teha universaalsed töövahendid, mis võimaldavad teha eestikeelsete tekstide emotsionaalsuse analüüsi nii leksikoni baasil kui masinõppe-meetodil, kus nad treenisid Naïve Bayes ja SVM (Support Vector Machine) klassifitseerijaid oma koostatud tekstikorpuste pealt (Pajupuu jt, 2016). Tulenevalt eesti keele rikkalikust morfoloogiast otsustasid nad mitte kasutada emotsioone näitavate sõnade algvorme, sest nii võib sõna emotsionaalsus muutuda vastupidiseks, näiteks „nautima“ (positiivne) ja „nautimata“ (negatiivne). Ka eemaldasid nad tekstidest morfoloogilised homonüümid ehk siis sõnakujud, kus erinevad sõnad erinevates vormides omavahel kirja pildis kattuvad, näiteks „tänavat“ (positiivne, „tänavat inimest“) ja tänavat (neutraalne, „tänavat ületavat inimest“) (Pajupuu jt, 2016).

Leksikoni koostamise käigus märkasid autorid, et hea tulemuse saab siis, kui kasutada kõige levinumaid emotsionaalsusega sõnu, mis oleksid erinevates tekstides monovalentsed ehk siis nende tähendus ei muutuks erinevates tekstides kontekstist sõltuvalt (Pajupuu jt, 2016). Sellist lähenemist toetas teadmine, et eesti keeles kirjutatud keerukates tekstides kasutatakse ainult 3000 sõna 75% teksti loomiseks (Kerge jt, 2014), ehk siis suure

tõenäosusega on nende 3000 sõna hulgas olevad emotsionaalsusega laetud sõnad ka enamikes tekstides, mida analüüsida (Pajupuu jt, 2016).

Leksikoni koostamisel võeti aluseks Eesti keele põhisõnavara sõnastikus² olevad 3015 sõna, millest määrati positiivse emotsiooniga 317 ja negatiivse emotsiooniga 322 sõna. Peale sõnastikus olevatele sõnadele antonüümide ja sünonüümide lisamist oli leksikonis 617 sõna positiivse emotsiooniga ja 730 sõna negatiivse emotsiooniga. Sõnaliigiti piiranguid ei tehtud ja sõnastikku lisati sõnadest erinevad sõnakujud, kokku 38628 sõna (Pajupuu jt, 2016).

Leksikoni-põhises meetodis tehti lisaks kolm eitustega seotud reeglit ning tulemuseks saadi, et keskeltläbi töötasid mõlemad meetodid 75% täpsusega. Leksikoni-põhine meetod ei andnud häid tulemusi kommentaaride emotsionaalsuse määramisel (58,3%), mida leksikoni loojad põhjendasid sellega, et kommentaarid on tekstidena lühemad, samas kui teksti emotsionaalsuse määramisel leksikoni abil oleks vaja, et teksti pikkus oleks vähemalt kaks omavahel sisuliselt seotud lauset. Keeleliselt rikkalike kultuuritekstide emotsionaalsuse määramisel oli leksikoni-põhise meetodi täpsus hea (78,2%), samas kui masinõppe-põhine meetod näitas kultuuritekstide emotsionaalsuse määramisel kõige väiksemat täpsust (Naïve Bayes 54,4% ja SVM 58,1%) (Pajupuu jt, 2016).

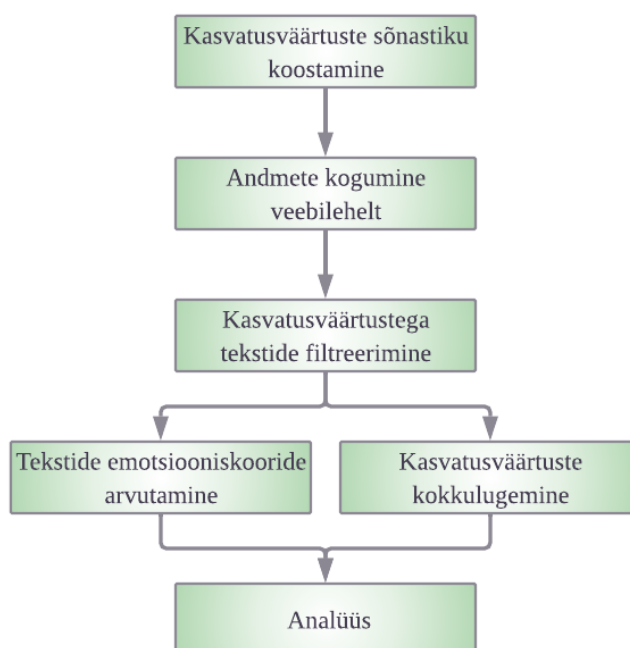
Kasvatuväärtused ja emotsioonianalüüs on tähtsal kohal järgmistes peatükkides käsitletavas metoodikas, mille abil filtreeritakse kasvatuväärtusi sisaldavad tekstid ja arvutatakse tekstide emotsiooniskoorid.

²<http://www.eki.ee/dict/psv/index.cgi?C06=en>

3. Tekstide filtreerimise metoodika

3.1 Metoodika kirjeldus

Magistritöös luuakse metoodika koos toimiva tarkvaralahendusega, mis võimaldab alla laadida andmeid Perekooli foorumist Koolilapse teema alt, otsida andmetest üles kasvatusväärtusi sisaldavad tekstid ja siis neid analüüsida, määrates ka tekstide emotsionaalsuse (joonis 3). Lisas 1 on juhend, kuidas seadistada tarkvara kasutamiseks vajalik keskkond.



Joonis 3. Magistritöö tarkvaralahenduse loomise tööprotsess

Metoodikas käiakse läbi alljärgnevad etapid:

1. Eesti keele töötlemise teegi EstNLTK tutvustamine;
2. Kasvatusväärtuse sõnastiku loomine;
3. Veebiämbliku tarkvara loomine ja andmete allalaadimine Perekooli foorumist;
4. Kasvatusväärtuste sõnastiku loomine;
5. Kasvatusväärtusi sisaldavate tekstide filtreerimine;
6. Kasvatusväärtusi sisaldavate andmete analüüs.

Kõikide etappide tarkvarakood asub GitHub'is³ ja iga etapi juures on ka link veebiaadressile, kus asub antud etapi kommenteeritud tarkvarakood.

³<https://github.com/Teet1970/Magistrit->

3.2 Uurimuse objekt ja hüpoteesid

3.2.1 Uurimuse objekt

Perekooli foorum (www.perekool.ee) loodi 1999. aastal Eesti Ämmaemandate Ühingu poolt ning Avatud Eesti Fondi toetusel eesmärgiga pakkuda tasuta ja professionaalset nõustamist kõigi perega seotud teemade osas. Aastal 2015 oli kodulehel 30500 registreeritud kasutajat ja vahemikus 80000-108000 unikaalset külastajat nädalas⁴. Ajavahemikul 01.01.2019 - 22.04.2019 oli veebilehel nädalas vahemikus 75000-95000 unikaalset külastajat (*Absolute Unique Users*)⁵, seega veebilehe külastatavus on küll kerges langustrendis, aga teiste sotsiaalvõrgustike kasvavat populaarsust arvestades on Perekooli foorum suutnud oma külastatavust hoida heal tasemel.

Perekooli foorumis saavad postitusi teha ainult registreeritud kasutajad ning soovi korral saavad nad teha seda anonüümselt, sellisel juhul on postituse autoriks märgitud kõikidel juhtudel „Kägu“.

Perekooli Koolilapse alamfoorum sai uurimuse objektiks valitud põhjusel, et esmaste vaatluste tulemusel selgus, et alamfoorumis on palju kasvatusväärtustega seotud teemasid ning tekstid on sotsiaalmeedia kontekstis pikad, seega on võimalik leida kasvatusväärtustega seotud informatsiooni.

3.2.2 Hüpoteesid

Magistritöös püstitati kaks hüpoteesi:

1. Tulviste, Mizera ja De Geer (2012) uurisid Eestis elavate emade kasvatusväärtusi ning nii eesti kui vene keelt emakeelena rääkivate emade eelistatuimad kasvatusväärtused olid traditsiooniliste kasvatusväärtuste hulka kuuluvad „heatahtlikkus“ ja „konformsus“. Püstitati hüpotees, et analüüsitava ajavahemiku jooksul (2012-2018) on autonoomsete kasvatusväärtuste osakaal võrreldes traditsiooniliste kasvatusväärtustega suurenenud, sest ühiskonnas ja meedias on autonoomsed kasvatusväärtused perioodi jooksul palju kõlapinda saanud.
2. Teine hüpotees lähtub eeldusest, et traditsioonilisi kasvatusväärtusi hindavad vanemad on rohkem mures laste käitumise pärast ning seetõttu on traditsioonilistel kasvatusväärtustel keskmised emotsiooniskoorid madalamad kui autonoomsetel kasvatusväärtustel.

⁴ <https://www.err.ee/548087/pealtnagija-ammaemandate-uhingu-suhtluskeskkonnast-perekool-on-saanud-eesti-oelaim-netifoorum>

⁵ <https://metrix.station.ee/site/510/www-perekool-ee>

3.3 Teek EstNLTK

EstNLTK⁶ on Pythoni programmeerimiskeeles kirjutatud eesti keele töötluste teek, mille loomisel on eeskujuks võetud keeletöötlusteegid TextBlob ja NLTK (Orasmaa jt, 2016). Teegi loomise põhjuseks oli asjaolu, et kuigi eesti keele töötlusteks oli loodud mitmeid erinevaid programme, oli nende kooskasutamine väga komplitseeritud, sest kasutati erinevaid programmeerimiskeeli, andmeformaate ja neil olid erinevad nõudmised arvuti tark- ja riistvarale. Tartu ülikooli andmeteadlaste poolt loodud teek EstNLTK ühendab erinevad olemasolevad tarkvarakomponendid ühte teeki nii, et neid saab kõiki koos kasutada (Orasmaa jt, 2016).

2019. aasta kevadel on täisfunktsionaalsena kasutamiseks EstNLTK versioon 1.4.1, samas on võimalik kasutada ka arendamisel olevat uuemat 1.6 beta versiooni.

Teegi kasutamiseks tuleb importida teek ja tekst muuta klassi *Text*'i objektiks, näiteks:

```
from estnltk import Text

lause = "Tänavu tuleb kevad teisiti."

tekst_objekt = Text(lause)
```

Text on standardse Pythoni sõnastiku alamklass koos täiendavate meetodite ja omadustega, mis on disainitud just keeletöötlust silmas pidades. Nii on ka näitena toodud *tekst_objekt* uus sõnastik. Sõnastiku kasutamisel baasandmeformaadina on mitmeid eeliseid – seda on lihtsam kontrollida ja parandada, on laiendatav, saab salvestada meta-andmeid ja serialiseerida JSON⁷ formaati.

Loodud uues sõnastikus on andmed salvestatud märgenduskihtidena. Iga märgenduskiht on omakorda sõnastik, mis lisaks kohustuslikele algus- ja lõpp-punkti atribuutidele võib salvestada erinevat vajalikku informatsiooni (Orasmaa jt, 2016).

EstNLTK põhifunktsioonid (Orasmaa jt, 2016)⁸:

1. Teksti jagamine väiksemateks osadeks – lõikudeks, lauseteks ja sõnadeks. Lõikudeks jagamisel võetakse aluseks tühja rea olemasolu tekstis.
2. Morfoloogiline analüüs. Võimaldab eesti keeles leida sõnade lemmad (algvormid), sõnaliigid, sufiksid, sõnalõpud ja sõnavormid. Kõigi sõnade puhul ei oska programm konteksti põhjal korrektset lemmat leida ja vigade vältimiseks annab mitu lemmat.
3. Õigekirja kontroll ja morfoloogiline süntees⁹ (morfoloogilise analüüsi pöördprotsess – võimaldab genereerida etteantud lemmale ja vormitunnustele vastava sõnavormi).

⁶<https://estnltk.github.io/estnltk/1.4.1/tutorials/text.html>

⁷<https://en.wikipedia.org/wiki/JSON>

⁸http://www.lrec-conf.org/proceedings/lrec2016/pdf/332_Paper.pdf

⁹<https://estnltk.github.io/estnltk/1.4.1/tutorials/text.html#morphological-synthesis>

4. Nimeüksuste tuvastamine (*named entity recognition* – NER). Võimaldab tekstist leida inimeste, organisatsioonide ja asukohtade nimed.
5. Ajaväljendite määramine. Võimaldab tekstis määrata ajaväljendid. Kui ajaväljend on relatiivse semantikaga (näiteks sõnad „eile“ ja „laupäeval“), siis võetakse kuupäeva määramisel aluseks programmi käivitamise kuupäev. Kui soovitakse muuta teksti loomise kuupäeva, muudetakse kuupäev atribuuti *creation_date* kasutades, kirjutades kuupäeva üle soovitud kuupäevaga.
6. Osalauseite leidmine. Võimaldab pikad ja keeruka ülesehitusega laused tükeldada väiksemateks osadeks. Suudab eraldada nii üksteisele järgnevaid osalauseid kui üksteise sees olevaid osalauseid (kiillauseid).
7. Verbiahelate tuvastamine. Leiab verbidega seotud ühe- ja mitmesõnalised üksused. Mitmesõnaliste üksuste puhul on võtme '*roots*' all olevad algvormid järjestatud sõnadevaheliste seoste alusel ja iga järgnev sõna ahelas on eelmise sõna süntaktiline alluv. Ahela eesotsas on tähendust mõjutavad verbid, näiteks modaalverb (näiteks „võima“, „tohtima“, „pidama“) või eituse abiverb ning ahela lõpus sisuverb, mis annab edasi väite põhisisu (Orasmaa ja Särg, 2017).
8. Grammatikate kirjutamise töövahend reeglipõhise info eraldamise jaoks. Reeglite koostamisel saab nõ ehitusklotsidena kasutada sõnaliike, lemmasid ja regulaaravaldisi (*regex*)¹⁰, kombineerides neid omavahel. Näiteks leia tekstist kohad, kus järjest on omadussõna ja nimisõna, mis on omavahel eraldatud tühikuga.

```
lause = Text („Täna on ilus ilm ja soe päike paistab!")
omadussõna = Postags('A','C','U')
nimisõna = Postags('S')
tühik = Regex('\s')
otsi_omadussõna_nimisõna = Concatenation(omadussõna, nimisõna, sep =
tühik)
print(otsi_omadussõna_nimisõna.get_matches(lause))
```

Tulemus:

```
[{'end': 16, 'matches': {}, 'text': 'ilus ilm', 'start': 8},
{'end': 29, 'matches': {}, 'text': 'soe päike', 'start': 20}]
```

Soovi korral võib reeglites ka täpsustada omadussõna, näiteks leia kohad, kus lauses üksteisele järgnevad omadussõna „ilus“ ja nimisõna, mis on omavahel eraldatud tühikuga.

```
ilus = Lemmas('ilus')
nimisõna = Postags('S')
tühik = Regex('\s')
otsi_ilus_nimisõna = Concatenation(ilus, nimisõna, sep = tühik)
print(otsi_omadussõna_nimisõna.get_matches(lause))
```

¹⁰<https://et.wikipedia.org/wiki/Regulaaravaldis>

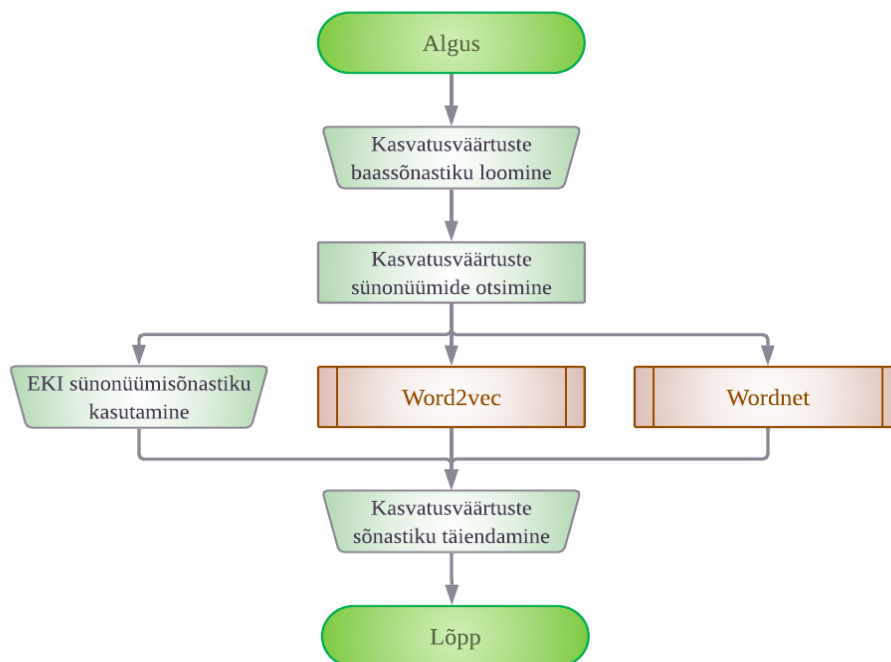
Tulemus:

```
[{'end': 16, 'matches': {}, 'text': 'ilus ilm', 'start': 8}]
```

3.4 Kasvatusväärtuste sõnastiku loomine

3.4.1 Kasvatusväärtuste baassõnastiku loomine

Leidmaks üles kasvatusväärtusi sisaldavad tekstid, tuli koostada kasvatusväärtuste sõnastik. Sõnastiku koostamiseks tehti antud tööetapist protsessijoonis (joonis 4).



Joonis 4. Kasvatusväärtuste sõnastiku loomine

Joonis 4 näitab, et sõnastiku koostamises on kaks suuremat etappi – esmalt tuleb koostada kasvatusväärtuste baassõnastik ja siis teises etapis tuleb leida kasvatusväärtustele võimalikult palju sünonüüme.

Kuna kasvatusväärtusi ei ole varem Eestis sellise metoodikaga *online*-meediatekstidest otsitud, siis puudus kasvatusväärtuste sõnastik. Baassõnastiku loomisel võeti aluseks Schwartzi individitasandi ja kultuuritasandi väärtusklastrid, mis on eesti keelde tõlgitud Veronika Kalmuse poolt (Kalmus, 1999). Baassõnastikku täiendati Tiia Tulviste ja Luule Mizera poolt koostatud ankeedis „Küsimustik lapsevanemale“ (lisa 3) leiduvate kasvatusväärtustega (Tulviste ja Mizera, 2010).

Kasvatuse väärtused jagati vastavalt Schwartzi modifitseeritud dünaamilise struktuuriga kasvatuse väärtuste jaotusele neljaks (joonis 1):

- saavutamine;
- loomingulisus;
- konformsus;
- heatahtlikkus/kohustumine.

Täiendavalt ühendati „saavutamine“ ja „loomingulises“ ühise nimetaja „autonoomia“ alla ning „konformsus“ ja „heatahtlikkus/kohustumine“ ühise nimetaja „traditsioonid“ alla.

3.4.2 Kasvatuse väärtuste sünonüümide otsimine

Teises etapis laiendati kasvatuse väärtuste sõnastikku olemasolevatele kasvatuse väärtuste sünonüümide otsimise kaudu. Kõik kasvatuse väärtused on sõnastikus algvormi kujul ehk lemmadena ning sõnaliikidest on leksikonis esindatud omadussõnad, nimisõnad ja tegusõnad (lisa 4). Sünonüümide otsimiseks kasutati kolme võimalust:

1. Eesti Keele Instituudi sünonüümisõnastik¹¹.
2. Eesti Wordnet¹², mis on integreeritud EstNLTk teeki ja kasutamiseks tuleb importida moodul ja edasiste otsingute kiirendamiseks korraga alla laadida kõik sünonüümihulgad, mis ongi sõnade sünonüümid. Eesti Wordnet on leksikaal-semantiline andmebaas, mis sisaldab informatsiooni mõistete ja nende vaheliste semantiliste suhete kohta. Wordneti põhiühik on sünonüümihulk (ka sünonüümihulk) ning võimalik on leida nii sõnade ülemmõisteid ehk hüperonüüme kui ka alamõisteid ehk hüponüüme (Orav, Zupping ja Vare, 2014). Näiteks sõna „koer“ ülemmõiste on „koerlane“ ja alamõiste on „õuekoer“. GitHubis¹³ on näidispäring, kuidas leida sünonüümid sõnale „viisakas“.
3. Word2Vec – suhteliselt vähest arvutusvõimsust vajav masinõppe algoritm, leidmaks tekstist seoseid sõnade vahel. On võimalik kasutada kahte mudelit - Continuous Bag-of-Words (CBOW) ja Skip-Gram. Algoritmiliselt on mudelid sarnased, välja arvatud see, et kui CBOW mudel leiab nõ sihtsõnu (*target-words*) kogu lähteandmestiku kontekstist, siis Skip-Gram mudelis on vastupidi ja leitakse lähteandmestiku konteksti sõnad sihtsõnadest (Mikolov, Chen, Corrado ja Dean, 2013).

Skip-Gram mudel on oluliselt parem just semantilises analüüsis, samas kui süntaktilise analüüsi täpsus on mudelitel sarnane. Skip-Gram'i mudeli treenimisel on olulised parameetrid dimensionaalsus ehk kui mitu vektorit on ühel sõnal teiste sõnade suunal ja aken (*window*) ehk kui mitut sõna enne ja pärast sõna analüüsitakse (maksimumpikkuseks on lause). Mida suurem dimensionaalsus ja mida

¹¹ <http://www.eki.ee/dict/sys/>

¹² <https://estnltk.github.io/estnltk/1.4.1/tutorials/text.html#estonian-wordnet>

¹³ https://github.com/Teet1970/Magistrit/blob/master/1_mag_Eesti_Wordnet.ipynb

suurem aken, seda parem tulemus, aga koos sellega kaasneb ka vajaliku arvutusvõimsuse arv. Algoritm arvutab sõnade omavahelisest kaugusest tulenevalt kaalud (*weights*) – mida lähemal sõnad üksteisele tekstis asuvad, seda suurem kaal (Mikolov jt, 2013). Sünonüümide leidmisel kasutataksegi kaalude andmeid – mida suurem on arv skaalal nullist üheni, seda sarnasemad on sõnade tähendused algoritmi arvates.

Sünonüümide leidmisel kasutati Eesti keele koondkorpuse¹⁴ (sisaldab 55 miljonit sõna ja 16 miljonit lauset) tekstide peal treenitud lemmatiseeritud Skip-Gram mudelit, mille dimensionaalsus on 200¹⁵. Erinevalt Estonian Wordnetist, kus on manuaalselt sõnadele lisatud nende sünonüümid, on Word2Vec antud kujul juhendamata masinõppe algoritm ja tulemustest leiab ka selliseid seoseid, mis on pärit mudeli treenimisel kasutatud korpusest, näiteks “sõnakuulekas” – “kommunaalpoliitik” ja “eesmärgikindlus” – “vargamäelik”.

Tulemuste võrdlemiseks treeniti mudel ka Perekooli foorumist allalaaditud tekstide pealt (dimensionaalsus 200, aken 5), aga tulemused jäid sünonüümide otsimise kontekstis palju tagasihoidlikumateks.

Kasvatusväärtustele sarnasuselt lähimad sõnad on nähtaval GitHubis¹⁶.

3.5 Veebiämbliku tarkvara loomine ja andmete allalaadimine Perekooli foorumist

Magistritöös oli üheks eesmärgiks ehitada veebiämblik, mis leiaks üles ja laadiks alla Perekooli foorumi “Laps” alamfoorumi „Koolilaps“¹⁷ all olevate postituste tekstid koos lisainfoga (kuupäev, teema, kasutaja id). Veebiämbliku mõiste ja kasutamismeetodid on kirjas lisas 2.

3.5.1 Tarkvara algoritmi koostamine

Tarkvara algoritmi (tööprotsessi) visualiseerimiseks tutvuti alamfoorumi „Koolilaps“ ülesehitusega ning selgus, et osade teemade osas oli nii palju postitusi, et nad paiknesid rohkem kui ühel leheküljel.

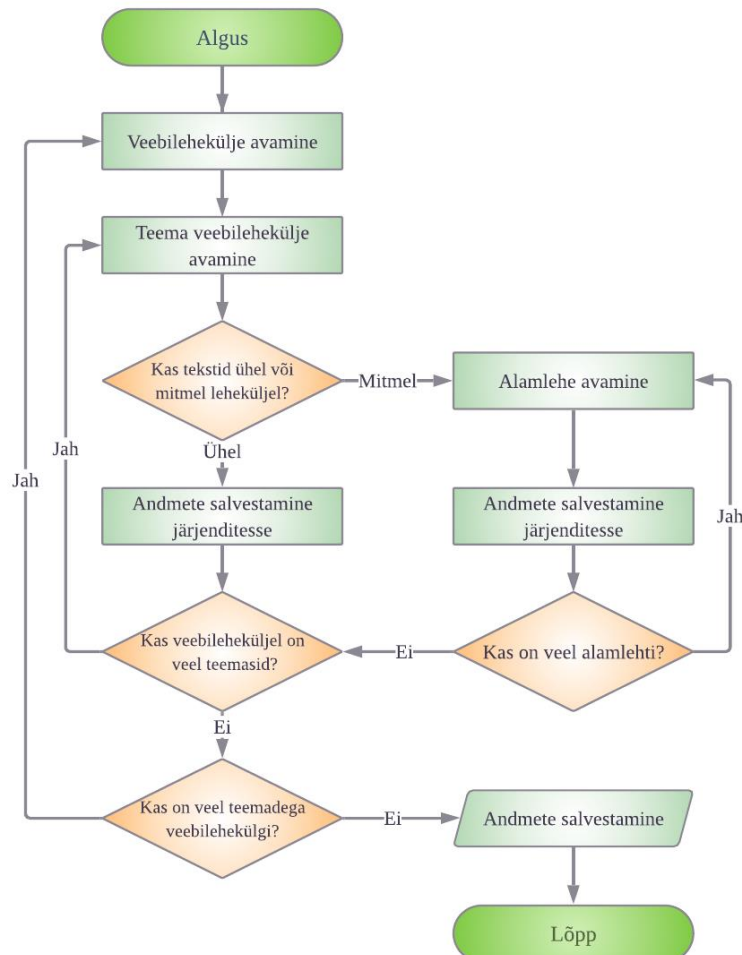
¹⁴<https://www.cl.ut.ee/korpused/segakorpus/index.php?lang=et>

¹⁵Treenitud mudelid on kasutamiseks mitteäriilistel eesmärkidel ja vabalt allalaaditavad veebiaadressilt <http://ats.cs.ut.ee/keeletehnoloogia/estnltk/word2vec/>

¹⁶https://github.com/Teet1970/Magistrit-/blob/master/2_mag_Word2vec.ipynb

¹⁷<https://foorum.perekool.ee/foorum/koolilaps/>

Joonisel 5 on näha, et osade teemade puhul toimub jagunemine – postitusi on konkreetse teema all nii palju, et nende jaoks on loodud täiendavad alamlehed ning programm peab suutma kõigilt alamlehtedelt koguda vajaminevad andmed.



Joonis 5. Perekooli Koolilapse foorumi veebiämbliku tööprotsess

3.5.2 Tarkvara loomine ja testimine

Kui tulevase programmi tööprotsess oli kaardistatud, tuli järgmises faasis kirjutada valmis programm, mis kavandatu ära teeks. Esmalt tutvuti veebilehe HTML (*Hypertext Markup Language*) ülesehitusega¹⁸ ning leiti, kus paikneb kahte liiki informatsioon:

- Info järgmiste ja viimaste lehekülgede kohta, et programm saaks järgmisele lehele liikuda, kui soovitud andmed on alla laaditud, ning lõpetaks töö, kui kõik andmed on alla laaditud. Näiteks info viimase lehekülje kohta on järgmine:

¹⁸ [view-source:https://foorum.perekool.ee/foorum/koolilaps/](https://foorum.perekool.ee/foorum/koolilaps/)

```
<li><a class='page-numbers'  
href='https://foorum.perekool.ee/foorum/koolilaps/page/263/'>263</a><  
/li>
```

Saadi teada, et kokku on antud teema all lehekülgi 263. Veebilehe ülesehitus oli sarnane juhtumitel, kus mingi teema juures on palju postitusi, mistõttu olid nad jaotatud mitmete alamlehtede vahel. Seal kasutati juba loodud koodi, mis oli paigutatud funktsioonide sisse.

- Postitustega seotud andmed – tekst, kuupäev, kasutaja id, teema. Teemade loeteluga veebilehelt saadi teema, konkreetse teema veebilehele liikudes saadi lisaks tekstid, kuupäevad ja postituste tegijate identifikaatori koodid.

Andmete parsimisel teeki *BeautifulSoup* kasutades märgati, et koos meid huvitavate andmetega tuleb kaasa ka palju ebaolulist infot. Andmete puhastamisel kasutati kahte meetodit:

- Ebasoovitava info kustutamiseks käsku *decompose()*, näiteks alljärgnev kood eemaldab *javascriptis* kirjutatud koodi:

```
for kood in soup("script"):  
    kood.decompose()
```

- Soovitava info leidmiseks kasutati regulaaravaldisi.

Veebiämbliku arhitektuuriks kujunes lõpus kaks üksteise sees asuvat *while*-tsüklit. Kuna sekundiline paus päringute vahel tegi programmi töö aeglasemaks, siis sai programmi testida, asendades viimase lehekülje leidmise funktsiooni ajutiselt meie poolt väiksema numbriga, näiteks:

```
while lehekylg <= 3:
```

3.5.3 Andmete allalaadimine

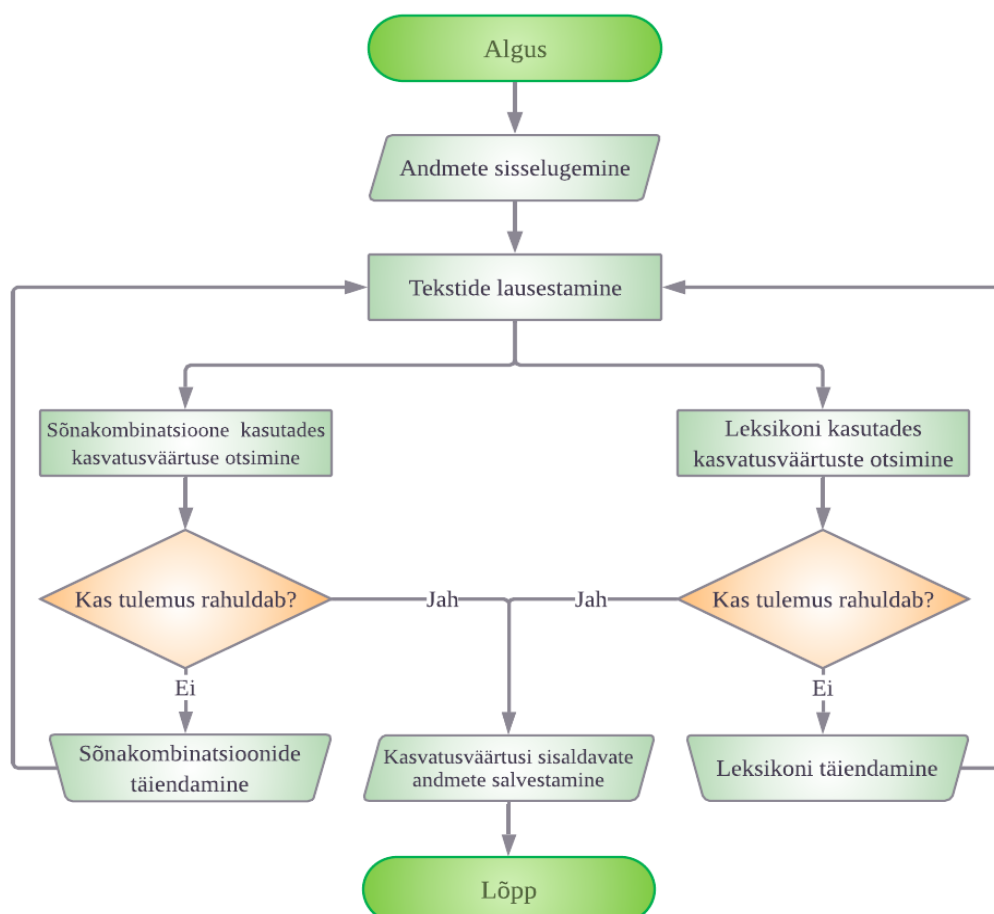
Kui testimine oli edukalt lõpule viidud, siis järgmine ja viimane samm oli andmete allalaadimine, mis viidi läbi 10. veebruaril 2019. aastal. Andmete allalaadimiseks ja teaduslikuks kasutamiseks andis Eesti Ämmaemandate Ühingu juhatuse nimel loa Marge Mahla. Kuna sekundiline paus päringute vahel aeglustas programmi tööd ja postitusi oli üle 158000, siis tekkis probleem, et peale viietunnist allalaadimist lõpetas väike häire internetiteenuses programmi töö. Selliste probleemide edasiseks vältimiseks otsustati andmed alla laadida kahes osas ja hiljem andmed omavahel liideti. Selleks leiti koodist viimase lehekülje arv (näiteks 263) ning tehti muudatused programmi muutujatesse.

Veebiämbliku tarkvara kood asub GitHubis¹⁹.

¹⁹https://github.com/Teet1970/Magistrit/blob/master/3_mag_veebiamblik.ipynb

3.6 Kasvatustähtuste sisaldavate tekstide filtreerimine

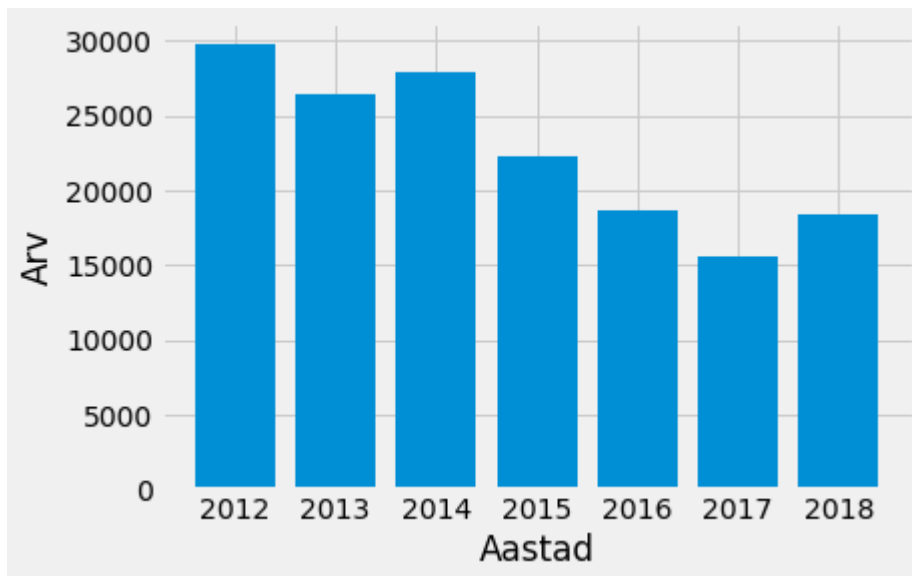
Kasvatustähtuste leksikon oli loodud ja järgmise sammuna oli vaja testida, kas leksikoni järgi leitakse kasvatustähtused soovitud kujul ja mida oleks vaja muuta nii kasvatustähtuste sõnastikus kui grammatikas, mille alusel leiame kasvatustähtusi sõnakombinatsioonide kujul, näiteks „sõna kuulama“. Joonis 6 näitab andmete filtreerimise protsessi, mis oli vaja läbi teha, enne kui saadi andmed salvestada.



Joonis 6. Kasvatustähtuste sisaldavate tekstide filtreerimine

Enne testimise alustamist vaadati, kuidas jagunevad tekstid aastate lõikes (kokku 158795 teksti). Jooniselt 7 on näha, et postituste arv oli langustrendis kuni aastani 2017, samas kui aastal 2018 pöördus see tõusule. Samal ajal Perekooli foorumi külastatavus ei ole langenud – kui aastal 2012 oli nädalas 61091-83502 unikaalset külastajat, siis aastal 2018 oli nädalas 78477-104105 unikaalset külastajat²⁰.

²⁰ <https://metrix.station.ee/site/510/www-perekool-ee>



Joonis 7. Perekooli foorumi Koolilapse alamfoorumi tekstide arv ajavahemikus 2012-2018

Mõned märkused koodi kirjutamise protsessi juurde (kood GitHubis²¹):

- Kuna kokku on tekste üle 150000, siis alustati testimist väiksema andmehulga peal (5000 teksti).
- Valikust jäid välja tekstid, kus on vähem kui 2 sõna.
- Kuna eesmärgiks oli eeskätt üles leida kasvatusväärtustega tekstid, siis selles etapis kasutati jaotust, kus kasvatusväärtused olid jaotatud kaheks – traditsioonilised ja autonoomsed.
- *Online*-tekstides on levinud lausete kokkukirjutamine ilma tühikuta. Kui laused olid jäetud tühikuteta eraldamata ja lõppesid punkti, hüüumärgi või küsimärgiga, siis tekitati sinna lausete vahele tühik.
- Lemmatiseeriti laused. Kui lemmatiseeritud sõna jäi analüüsi järel mitmeks (näiteks „kuulama|kuulanud“), siis võrreldi kõiki EstNLTK poolt pakutud algvormides sõnu kasvatusväärtuste sõnastikus leiduvatega.
- EstNLTK grammatika kirjutamise töövahendiga loodi sõnapaarid „sõna kuulama“, „kuulama sõna“, „käsku täitma“, „kohustusi täitma“.
- Kasvatusväärtusi sisaldavad laused paigutati sõnastikesse, kus võtmeteks olid kasvatusväärtused ja väärtusteks laused.

Esmane tulemuste analüüs näitas, et probleemid tekkisid sõnadega, mida eesti keeles kasutatakse ka teistsuguses kontekstis, nii esinesid omadussõnad „aus“, „viisakas“ ja „korralik“ sagedasti hoopis riietuse iseloomustamise kontekstis, millel puudus kasvatusväärtustega side. Olukorra parandamiseks testiti neid omadussõnasid koos erinevate nimisõnadega ning lõpptulemusena loodi EstNLTK grammatika kirjutamise

²¹https://github.com/Teet1970/Magistrit/blob/master/4_mag_kasvatusvaartused_faili.ipynb

töövahendiga uued grammatikareeglid, kus omadussõnu arvestati edaspidi kasvatusväärtuste kontekstis ainult siis, kui neile järgnesid nimisõnad „käitumine“, „laps“, „õpilane“, „koolilaps“, „poiss“, „tüdruk“, „komme“, „kasvatus“, „kasvatamine“.

Järgmisena pandi programm tööle kogu andmefaili ulatuses. Muudetud grammatika kasutamine tekitas siiski ka mittesooitud tulemusi, aga neid oli mõni üksik. Näiteks leidis grammatika üles sõnapaariid „korralik lapsevanem“ ja „kuulame sõnatult“.



Joonis 8. Kasvatusväärtuste sõnapilv (tegemisel on kasutatud näidiskoodi²²)

Joonisel 8 on sõnapilve kujul näha leitud kasvatusväärtused – mida suurem kiri, seda rohkem kordi on vastav kasvatusväärtus tekstis esinenud. Traditsioonilised kasvatusväärtused on rohelised ja autonoomsed kasvatusväärtused oranžid sõnad. Traditsioonilistest kasvatusväärtustest esines kõige rohkem sõnasid „austama“ (328), „kohusetunne“ (256) ja „distsipliin“ (238), autonoomsetest kasvatusväärtustest leidis kõige rohkem sõnu „iseseisev“ (711), „enesekindlus“ (408) ja „loovus“ (307).

²² https://amueller.github.io/word_cloud/auto_examples/colored_by_group.html

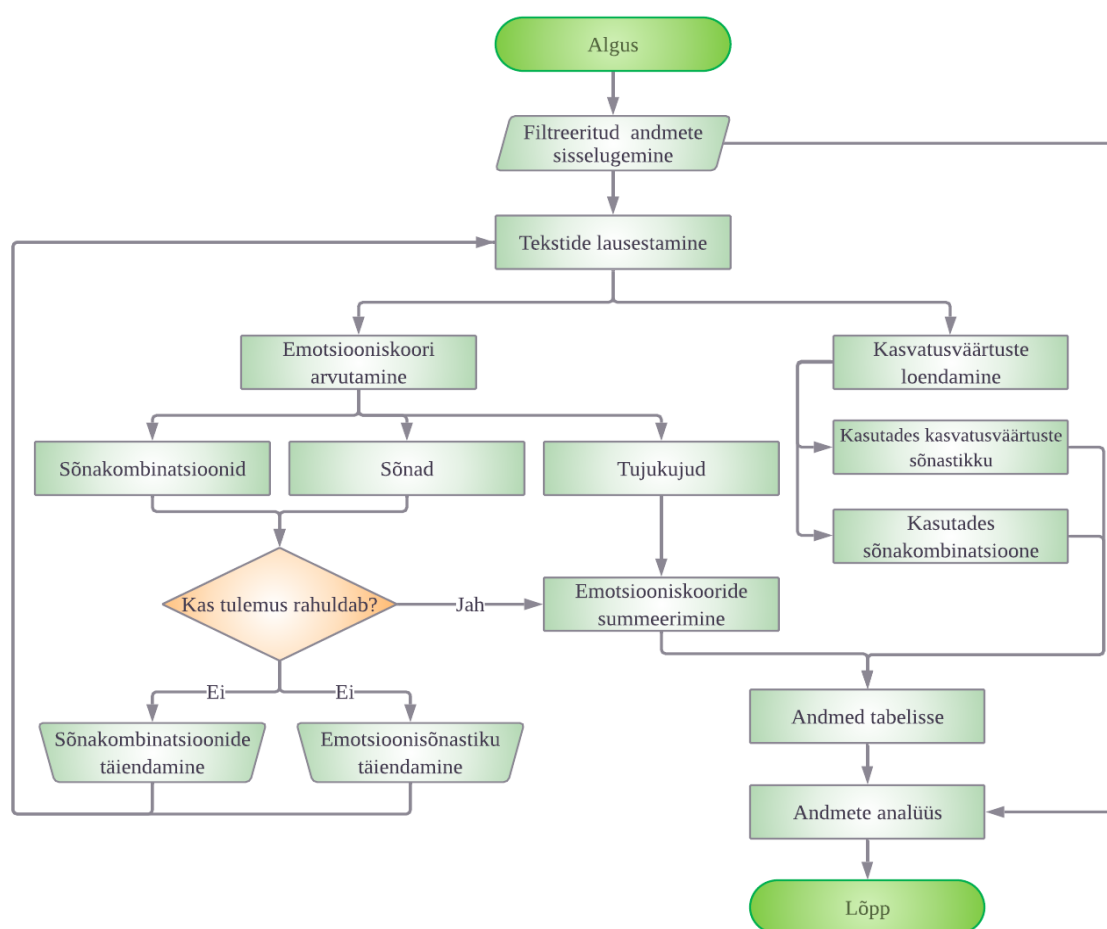
Kuna programmi töö fookus oli siiani kasvatusväärtuste leidmisel, aga neid võis ühes tekstis olla mitu, siis olid järjendis andmed dubleeritud ning neid tuli muuta nii, et järgi jääksid ainult unikaalsed andmerealad. Et hulka saab kasutada ainult mittemuteeritavate muutujatega, aga andmed olid järjendites, mis on muteeritavad, siis tuli andmed vahepeal muuta mittemuteeritavateks ennikuteks (*tuple*), siis teha hulgaks ning kõige lõpuks muuta järjenditeks tagasi. Viimase sammuna lisati iga andmerealade algusesse järjekorranumber ning kasvatusväärtusi sisaldavad tekstid salvestati faili.

Kui kasvatusväärtusi sisaldavad tekstid olid filtreeritud, tuli metoodikas edasi liikuda järgmisesse etappi, kus arvutatakse tekstide emotsiooniskoorid ning analüüsitakse andmeid.

4. Analüüs ja tulemused

4.1 Emotsiooniskooride arvutamine

Jooniselt 9 on näha, kuidas toimub emotsiooniskooride arvutamise ja kasvatusväärtuste kokkulugemise protsess.



Joonis 9. Tekstide emotsiooniskooride arvutamine ja kasvatusväärtuste kokkulugemine

Tekstide emotsiooniskooride arvutamisel võeti aluseks Pajupuu jt koostatud emotsioonileksikon koos kolme reeglina (Pajupuu jt, 2016):

1. Kui emotsioonisõnastikus olevale negatiivse emotsionaalsusega sõnale eelnevad eitused („ei“, „ega“, „pole“), siis muutub tegusõna emotsionaalsus positiivseks, näiteks „ei varastanud“.
2. Kui emotsioonisõnastikus olevale positiivse emotsionaalsusega sõnale eelnevad eitused („ei“, „ega“, „pole“), siis muutub tegusõna emotsionaalsus negatiivseks, näiteks „ei meeldi“.
3. Kui neutraalse tegusõna ees on eitus, muutub tegusõna emotsionaalsus negatiivseks.

Reeglite kasutamiseks tuli leksikonist (üle 39000 rea) välja otsida kõik emotsionaalsusega tegusõnad.

Online-meediatekstides on olulised kirjutaja emotsionaalsuse edasiandjad tujukujud ehk emotikonid. Kuna tujukuju leidmine Unicode koodiga eeldab korrektset kirjaipilti, aga meediatekstides on palju hooletusvigu just lauselõppudes, siis otsiti tujukujusid kahel viisil - Unicode koodi järgi ja tähemärkide kombinatsioonide järgi. Erinevate tujukujude leidmiseks vaadati läbi kõik tekstid ning kokku leiti kuus erinevat positiivset tujukuju tüüpi ja üks negatiivne tüüp. Kokku leiti 762 positiivset ja 51 negatiivset emotsionaalsust väljendavat tujukuju. Tabelist 1 on näha, et ühes tekstis oli kõige rohkem 5 positiivset ja 4 negatiivset emotsiooni väljendavat tujukuju.

Tabel 1. Tujukujude (emotikonide) arv ühes tekstis

Tujukujude arv tekstis	1	2	3	4	5
Positiivsed tujukujud	439	88	31	11	2
Negatiivsed tujukujud	37	3	0	2	0

Testimise käigus selgus, et kasvatusväärtusi sisaldavates tekstides on palju kodu ja perega seotud sõnu, mis emotsioonileksikonis olid kui positiivse emotsionaalsusega sõnad, aga antud kontekstis oli tegemist neutraalsete sõnadega, mis põhjustasid valepositiivsust, näiteks sõnapaar „õnnelik lapsepõlv“ annaks emotsionaalsusskoorile juurde 2 punkti, samas „õnnetu lapsepõlv“ skoor oleks 0. Nii eemaldati emotsioonisõnastikust sõnad „ema“, „isa“, „pere“, „kodu“, „eesti“, „vanaema“, „vanaisa“, „magustoit“, „moos“, „lapsepõlv“, „noored“, „lapselaps“, „laps“. Eemaldada tuli ka sõna „mees“, mis oli emotsioonisõnastikus kui sõna „mesi“ seesütlevas käändes vorm, aga tekstides sisaldus valdavalt sõna „mees“ teises tähenduses.

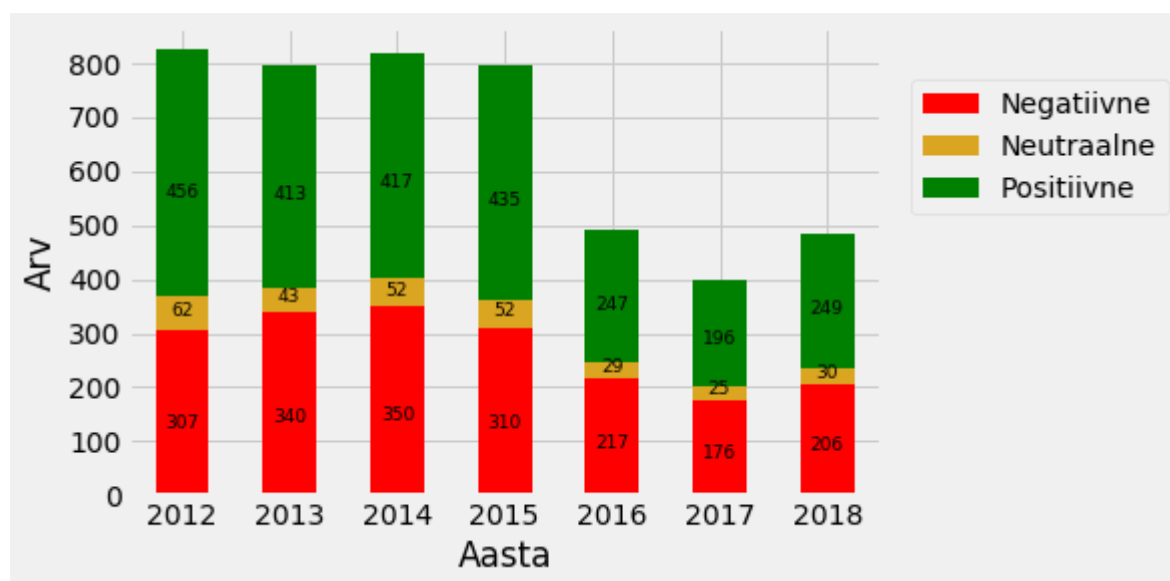
Probleem tekkis ka emotsioonisõnastikus oleva sõnaga „pole“, mis emotsioonisõnastikus oli negatiivse väärtusega, aga tekstides esines sõna tihtipeale väljendites, mis ei mõjutanud teksti emotsionaalsust, näiteks „pole oluline“. Lahendusena lisati reeglitesse sõna „pole“ tegusõna tähenduse muutjana (näiteks „pole varastanud“) ja eemaldati see sõna emotsioonisõnastikust.

Emotsioonisõnastikus on positiivse emotsionaalsusega sõnad skooriga 1 ja negatiivsed skooriga -1. Kuna Pajupuu jt (2016) polnud teadusartiklis välja toodud reeglite puhul skooride määramise meetodikat, siis alguses sai lähtutud eeldusest, et positiivsed sõnakombinatsioonid lisavad skoorile 1 ja negatiivsed sõnakombinatsioonid lahutavad skoorist 1 maha. Antud juhul tekkis probleem kolmanda reeglga, kus eituse eelnemine neutraalsetele tegusõnadele muudab nad negatiivse emotsionaalsusega sõnadeks, samas paljudel juhtudel see nii ei ole, näiteks lauses „ma ei soovinud meie kirglikku suudlust lõpetada“ ei muuda eituse lause emotsionaalsust negatiivsemaks, vaid see jääb neutraalseks.

Olukorra lahenduseks tehti täiendus, et kui neutraalsele tegusõnale eelneb eitus, väheneb teksti emotsionaalsusskoor 0,5 punkti võrra (kood GitHub'is²³).

4.2 Tekstide arv, emotsiooniskoorid ja kirjutamise aeg

Kokku filtreeriti 158795 tekstist välja 4612 kasvatusväärtusi sisaldavat teksti (2,9%). Jooniselt 10 näeme, et kasvatusväärtusi sisaldavate tekstide arv tegi kiire languse aastal 2016, samas kui analüüsi aluseks olevate tekstide koguarv (joonis 7) langes järjest aastani 2017 ning aastal 2018 kasvatusväärtusi sisaldavate tekstide arv taas kasvas.



Joonis 10. Kasvatusväärtustega tekstide jaotus aastate lõikes ja emotsiooniskoori järgi

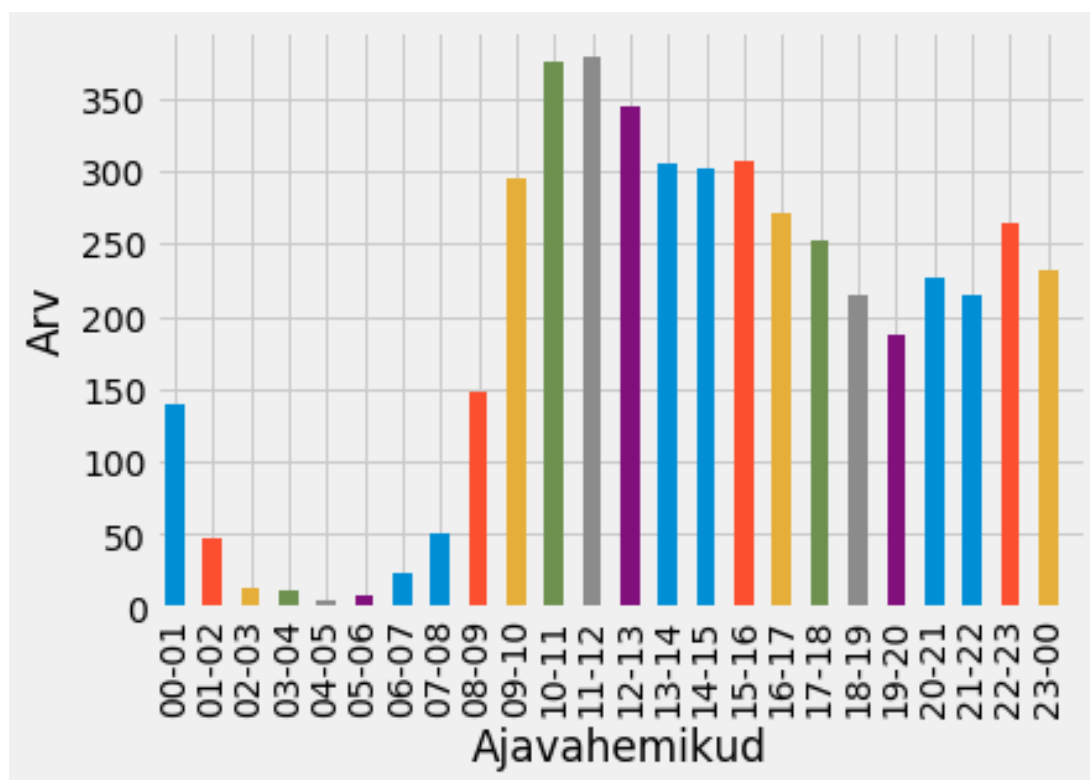
Teksti emotsionaalsus määrati 4576 tekstis (99,2%), emotsiooniskoor jäi nulli 257 tekstis, positiivse emotsiooniskooriga tekste oli 2413 ja negatiivse emotsiooniskooriga tekste 1906. Emotsionaalsus jäi määramata 36 tekstis, sest seal polnud ühtegi emotsionaalsusega tekstiühikut.

Tekstide jaotumisel emotsiooniskoori järgi näeme, et kõigil analüüsitud aastatel oli positiivse emotsiooniskooriga tekste rohkem kui negatiivse emotsiooniskooriga tekste, aga peale kasvatusväärtustega tekstide arvu arvestatavat langust 2016. aastal see suhtarv vähenes.

Saamaks ülevaadet, mis kellaaegadel kõige rohkem tekste on aastate jooksul loodud, jaotati ööpäev kolmeks – jaotus „öö“ hõlmab ajavahemikku 00.00-07.59, „päev“ hõlmab ajavahemikku 08.00-15.59 ja „õhtu“ hõlmab ajavahemikku 16.00-23.59.

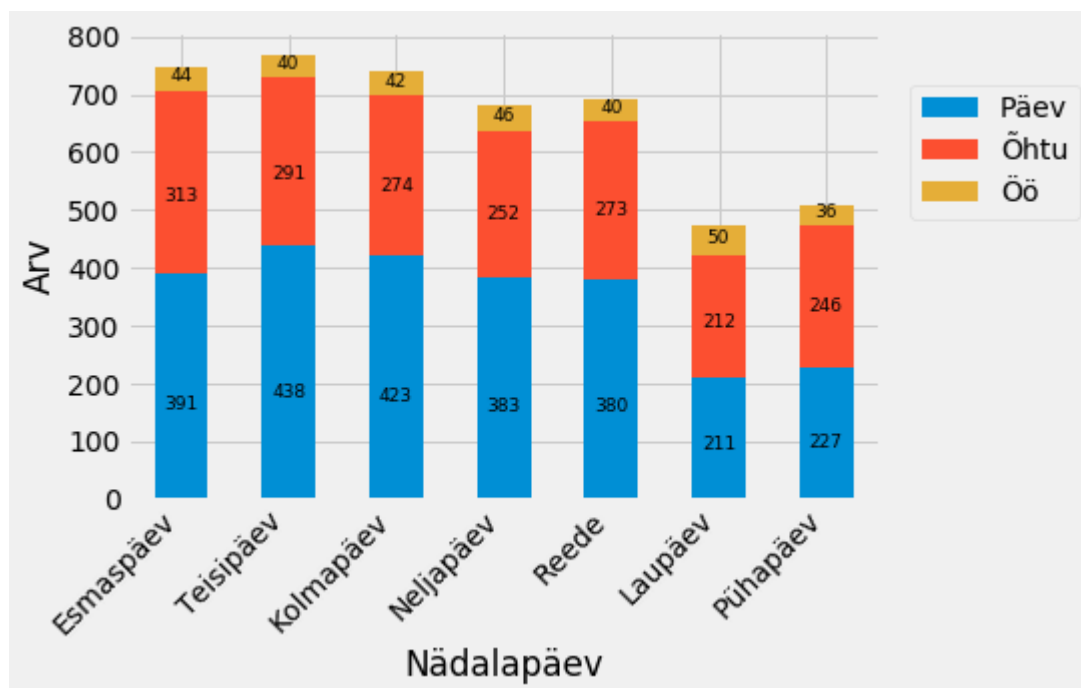
Jooniselt 11 nähtub, et kõige suurem osa kasvatusväärtusi sisaldavatest tekstidest on kirjutatud päevasel ajal, ajavahemikus 9.00 -15.59.

²³ https://github.com/Teet1970/Magistrit-/blob/master/5_mag_emotsiooniskoor_tabelisse.ipynb



Joonis 11. Tekstide kirjutamine tundide lõikes.

Võib-eeldada, et päevasel ajal kirjutatud tekstide arvud on suured seetõttu, et palju kasvatusväärtusi sisaldavaid tekste on kirjutatud nädalavahetustel päevasel ajal.



Joonis 12. Kasvatusväärtustega tekstide jaotus nädalapäevade lõikes

Joonist 12 vaadates võib näha, et nädalavahetustel kirjutatud tekstide arv on oluliselt väiksem, eriti just päevasel ajal. Seega kõige rohkem kasvatusväärtusi sisaldavaid tekste on kirjutatud nädala sees töö ajal.

4.3 Tekstide pikkus ja keskmine emotsiooniskoor

Sotsiaalmeedias avaldatud tekstide pikkused erinevad, sõltudes ka keskkonna poolt pandud piirangutest. Näiteks kuigi aastal 2019 on Twitteri säutsu maksimumpikkuseks 280 tähemärki, oli Sarah Perez'i andmetel²⁴ aastal 2018 ainult 5% säutsudest pikemad kui 190 tähemärki. Perekooli foorumist allalaaditud kasvatusväärtusi sisaldavate tekstide keskmine pikkus oli 145 sõna (sh. tujukujud) ja mediaan 115 sõna, keskmiselt oli ühes tekstis 960 tähemärki, 10,2 lauset ja ühes lauses oli keskmiselt 14,7 sõna, kusjuures sõnade arv lauses varieerus aastate jooksul vähe. Kasvatusväärtusi sisaldavate tekstide pikkus on oluline eeskätt seetõttu, et Pajupuu jt järgi on vaja teksti emotsionaalsuse paremaks määramiseks teksti, mille miinimumpikkus oleks kaks omavahel sisuliselt seotud lauset (Pajupuu jt, 2016).



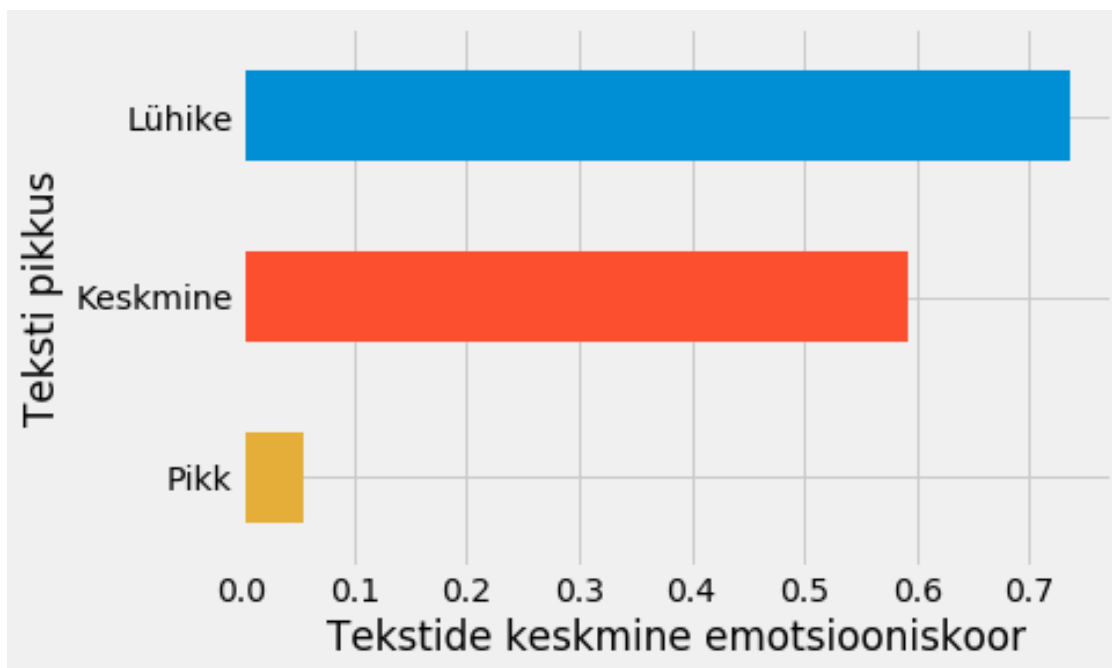
Joonis 13. Tekstide sõnade arvu I, II (mediaan), III kvartiil ja keskmine (valge ring) pikkus

Boxplot-tüüpi²⁵ joonisel 13 on näha teksti pikkusega seotud andmete muutust aastate lõikes. Kõikidel aastatel on tekstide keskmine pikkus suurem kui mediaan ehk siis tekstide hajuvus pikemate tekstide suunas on suurem. Kõige rohkem sõnu (3062) oli tekstis, mis osutus ajalehe Sirp artikliks (09.09.2015), mis käsitles eliitkoolide temaatikat ning mis oli täismahus kopeeritud Perekooli foorumisse.

²⁴ <https://techcrunch.com/2018/10/30/twitters-doubling-of-character-count-from-140-to-280-had-little-impact-on-length-of-tweets/>

²⁵ <https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51>

Vaadati ka tekstide emotsionaalsust tekstide pikkustega võrreldes. Selleks jagati tekstid kolme võrdsesse osasse – lühikesed (kuni 82 sõna), keskmised (83-155 sõna) ja pikad (156 ja rohkem sõna).



Joonis 14. Tekstide pikkuse keskmine emotsiooniskoor (emotsiooniskooride summa jagatuna tekstide arvuga).

Jooniselt 14 on näha, et pikkade tekstide keskmised emotsiooniskoorid on madalamad kui keskmise ja lühikese pikkusega tekstidel.

4.4 Kasvatuväärtuste arv

Kasvatuväärtusi loetleti tekstides kokku 5605 (tabel 2) ja keskmiselt sisaldas üks filtreeritud tekst 1,22 kasvatuväärtust.

Tabel 2. Kasvatuväärtuste koguarv

Kasvatuväärtused	Arv
Loomingulisus	2370
Saavutamine	469
Konformsus	1631
Heatahtlikkus/Kohustumine	1135
Kokku	5605

Kui kokkuvõttes oli autonoomseid (2839) ja traditsioonilisi (2766) kasvatuväärtusi samas suurusjärgus, siis suur erinevus oli autonoomsete kasvatuväärtuse alajaotuses - loominguilisusega seotud kasvatuväärtusi (2370) ja saavutamise seotud kasvatuväärtusi

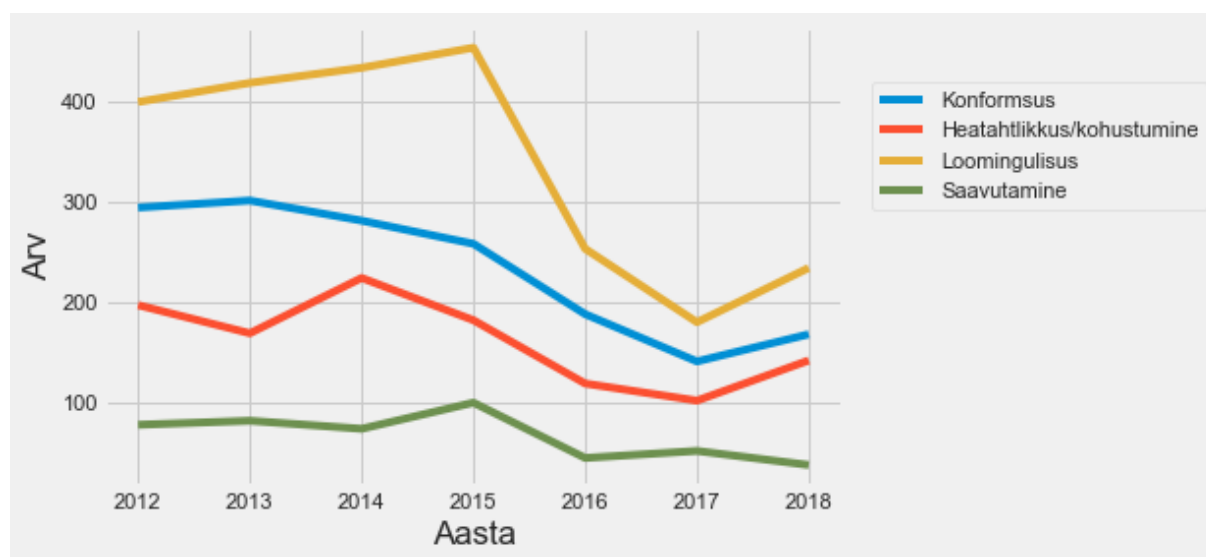
(469). Seda võib tõlgendada kui selget indikatsiooni, et saavutamise seotud kasvatusväärtused (näiteks „sihikindlus“, „ettevõtlikkus“, „järjekindlus“) ei ole võrreldes teiste kasvatusväärtustega nii olulised Perekooli foorumi tekstide kirjutajate jaoks.

Ühest tekstist leiti maksimaalselt (tabel 3) kümme loomingulisusega seotud, kuus konformsusega seotud, viis heatahtlikkuse/kohustumisega seotud ja kolm saavutamise seotud kasvatusväärtust. Valdavas enamuses tekstides (4302) oli siiski üks kasvatusväärtus või kaks (459) kasvatusväärtust.

Tabel 3. Kasvatusväärtuste arv ühes tekstis

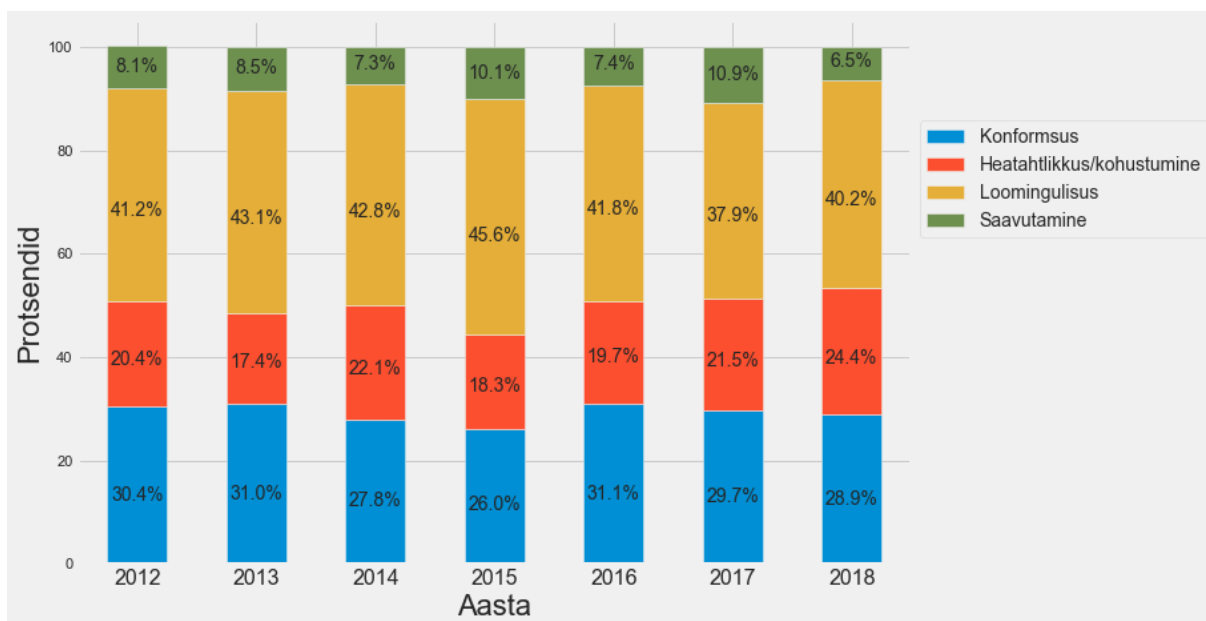
Kasvatusväärtuse arv	1	2	3	4	5	6	7	10
Loomingulisus	1753	209	41	8	3	2	1	1
Saavutamine	427	18	2	0	0	0	0	0
Konformsus	1232	139	24	8	1	2	0	0
Heatahtlikkus/Kohustumine	890	93	14	3	1	0	0	0
Kokku	4302	459	81	19	5	4	1	1

Jooniselt 15 nähtub, et kuigi loomingulisusega seotud kasvatusväärtuste osas on peale aastat 2015 olnud kõige järsem langus, on läbi aastate on kasvatusväärtuste jagunemise proportsioonid jäänud samaks.



Joonis 15. Kasvatusväärtuste jaotumine tekstides

Kuna aastate lõikes on ka analüüsitavate tekstide arv muutunud, siis parema ülevaate saamiseks vaadati kasvatusväärtuste suhtelist jaotumist aastate lõikes. Jooniselt 16 nähtub, et kuigi aastate jooksul on teatud kõikumisi, on kasvatusväärtuste jaotus jäänud samasse suurusjärku - kui 2012. aastal oli autonoomseid kasvatusväärtusi 49,3% loetud kasvatusväärtuste koguarvust, siis aastal 2015 küll korra kasv autonoomsete kasvatusväärtuste protsent 55,7 peale, aga hiljem langes taas ja viimasel uuritud aastal oli see 46,7%.



Joonis 16. Kasvatusväärtuste protsentuaalne jaotumine aastate lõikes

Seega ei leidnud kinnitust hüpotees, et autonoomsete kasvatusväärtuste suhteline esinemissagedus võrreldes traditsiooniliste kasvatusväärtustega on uuritava ajavahemikul kasvanud.

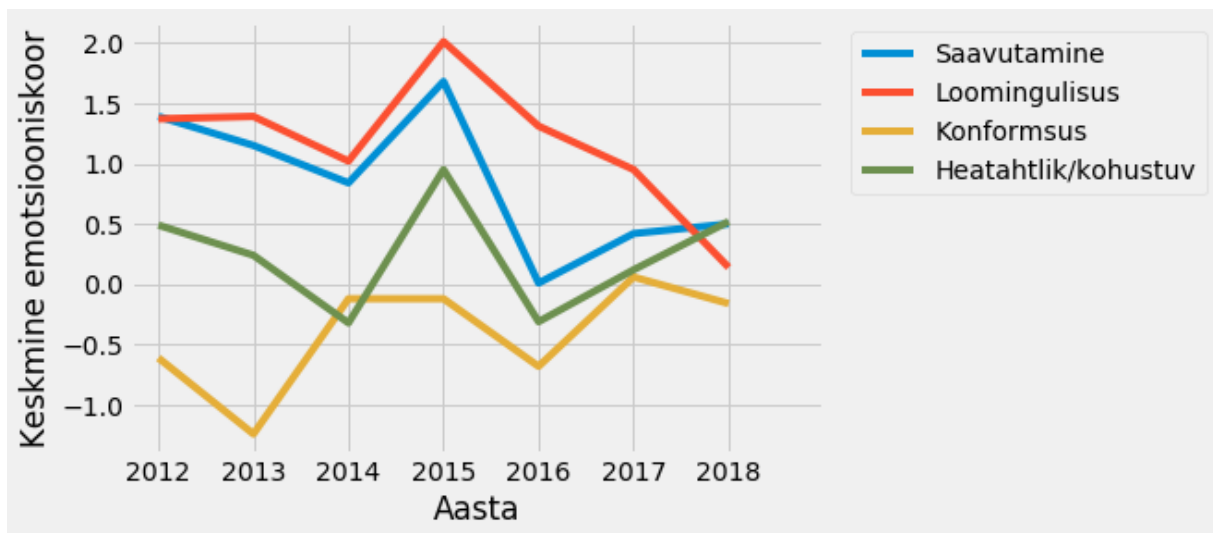
4.5 Kasvatusväärtuste väärtusgruppide ja väärtusjaotuse keskmine emotsiooniskoor

Kasvatusväärtusi analüüsiti kasvatusväärtuste keskmist emotsiooniskoori aluseks võttes kahel viisil:

- Väärtusgruppide järgi - autonoomsete kasvatusväärtuste alamjaotused „Saavutamine“ ja „Loomingulisus“ ning traditsiooniliste kasvatusväärtuste alamjaotused „Konformsus“ ja „Heatahtlikkus/kohustumine“.
- Väärtusjaotuse järgi - autonoomse kasvatusväärtuse leidmisel liideti kasvatusväärtuste summale number üks ja traditsiooniliste väärtuste leidmisel lahutati number üks. Kokku loodi nii viis kategooriat, kus „Võrdne“ tähendas sama arvu traditsiooniliste ja autonoomsete kasvatusväärtuste leidumist ühes tekstis, „Traditsiooniline“ tähendas tekstis ühe enama traditsioonilise kasvatusväärtuse leidumist võrreldes autonoomsete kasvatusväärtustega, „Traditsiooniline 2+“ tähendas tekstis kahe või enama traditsioonilise kasvatusväärtuse leidumist võrreldes autonoomsete kasvatusväärtustega, „Autonoomne“ tähendas tekstis ühe enama autonoomse kasvatusväärtuse leidumist võrreldes traditsiooniliste kasvatusväärtustega ja „Autonoomne 2+“ tähendas tekstis kahe või enama

autonoomse kasvatusväärtuse leidumist võrreldes traditsiooniliste kasvatusväärtustega.

Kasvatusväärtuste väärtusgruppide keskmine emotsiooniskoor arvatati erinevalt tekstide keskmisest emotsiooniskoorist, kus jagati keskmise saamiseks tekstide emotsiooniskooride summa tekstide arvuga. Autonoomsete kasvatusväärtuste alamjaotuste „Saavutamine“ ja „Loomingulisus“ ning traditsiooniliste kasvatusväärtuste alamjaotuse „Konformsus“ ja „Heatahtlikkus/kohustumine“ keskmise emotsiooniskoori arvutamiseks anti kõigile tekstis leiduvatele erinevatele kasvatusväärtuste alamjaotustele sama emotsiooniskoori väärtus, näiteks kui teksti emotsiooniskoor oli -10, siis said sama väärtuse (-10) nii tekstis leiduvad „saavutamise“ kui ka „loomingulisuse“ kasvatusväärtuste grupp ning kui mingi grupi kasvatusväärtusi oli tekstis rohkem kui üks, siis tulemus sellest ei muutunud. Kasvatusväärtuste keskmine emotsiooniskoor saadi kasvatusväärtuste alamjaotuse kasvatusväärtuste emotsiooniskooride summa jagamisel seda liiki kasvatusväärtusi sisaldavate tekstide arvuga.



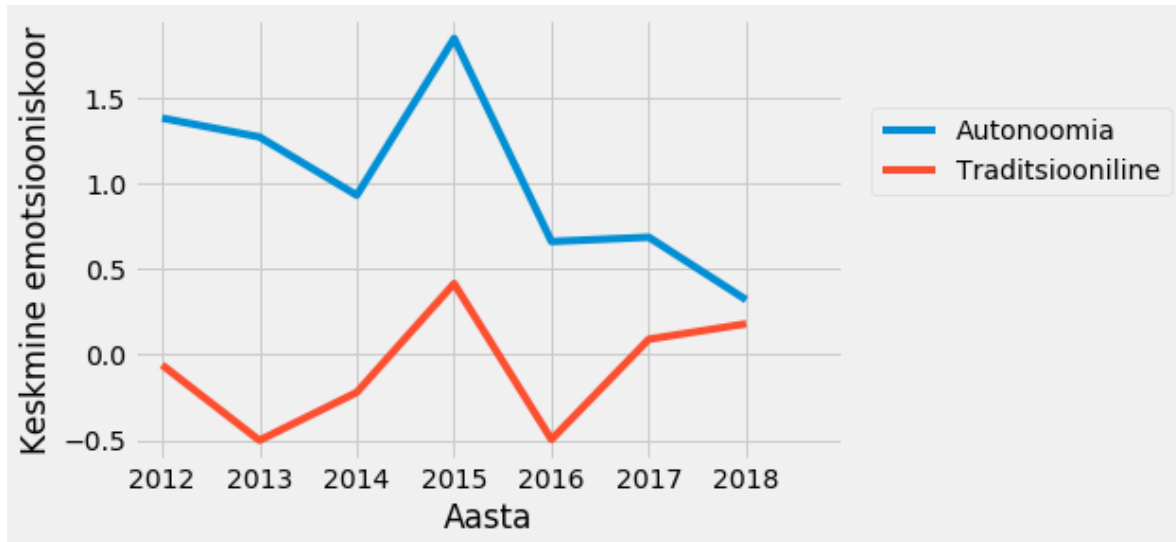
Joonis 17. Kasvatusväärtuste keskmine emotsiooniskoor

Joonisel 17 on näha, kuidas jagunevad autonoomsete kasvatusväärtuste alamjaotused „Saavutamine“ ja „Loomingulisus“ ning traditsiooniliste kasvatusväärtuste alamjaotused „Konformsus“ ja „Heatahtlikkus/kohustumine“. Kasvatusväärtuste grupid „Saavutamine“ ja „Loomingulisus“ olid ajavahemikul 2012-2015 sarnaste emotsiooniskooridega ning eristusid selgelt ülejäänud väärtusgruppidest. Kõige madalam emotsiooniskoor oli väärtusgrupil „Konformsus“. Peale 2015. aastat aga erinevus väheneb kiirelt kasvatusväärtuste gruppide vahel ning aastal 2018 on traditsiooniliste kasvatusväärtuste alamgrupi „Heatahtlik/kohustumine“ emotsionaalsusskoor tõusnud mööda kasvatusväärtuste grupist „Loomingulisus“ ja on samas suurusjärgus kui kasvatusväärtuste grupil „Saavutamine“.

Jooniselt 18 nähtava keskmine kasvatusväärtuse emotsiooniskoori arvutamiseks liideti joonisel 17 olevate väärtusgruppide „Saavutamine“ ja „Loomingulisus“ ning „Konformsus“ ja

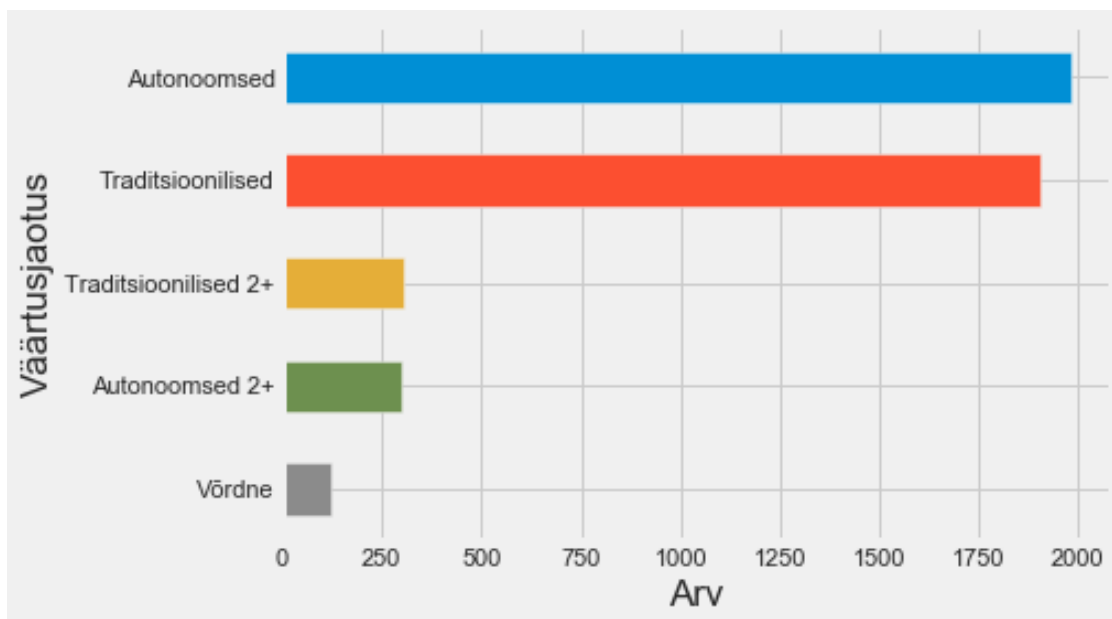
„Heatahtlikkus/Kohustumine keskmised kasvatusväärtuste emotsiooniskoorid ja jagati kahega.

Metoodika osas püstitati hüpotees, et traditsiooniliste kasvatusväärtuste keskmine emotsiooniskoor on madalam kui autonoomsetel kasvatusväärtustel.



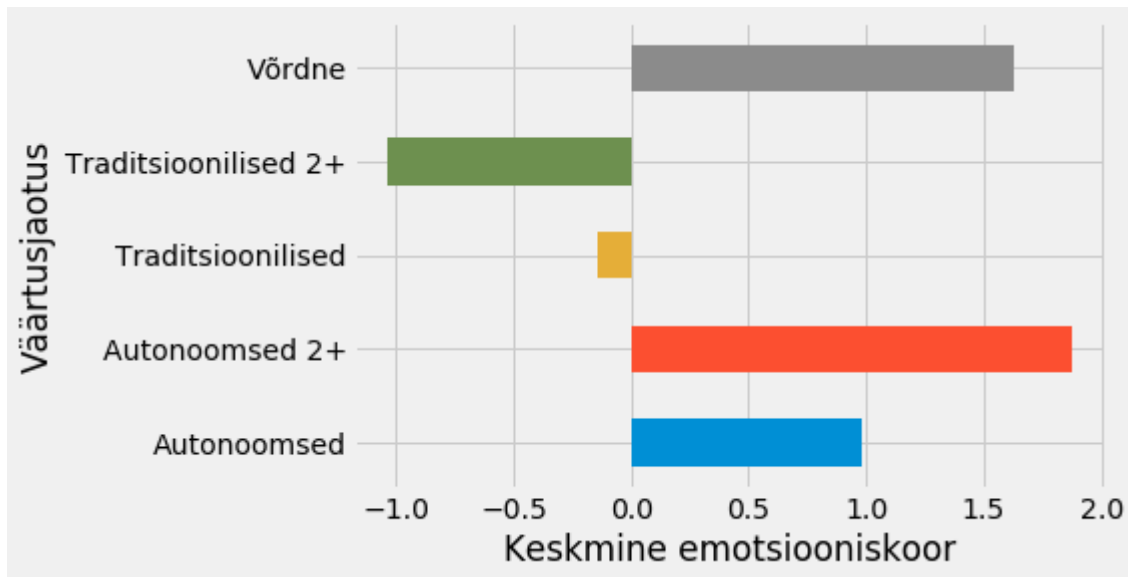
Joonis 18. Keskmine emotsiooniskoor traditsioonilistel ja autonoomsetel kasvatusväärtustel

Jooniselt 18 nähtub, et hüpotees leidis kinnitust, samas on viimastel aastatel emotsiooniskoori vahe vähenenud ning seda eeskätt seetõttu, et autonoomsete kasvatusväärtuste emotsiooniskoor on olnud langustrendis, samas kui traditsiooniliste kasvatusväärtuste emotsiooniskoor on aastate jooksul tõusnud.



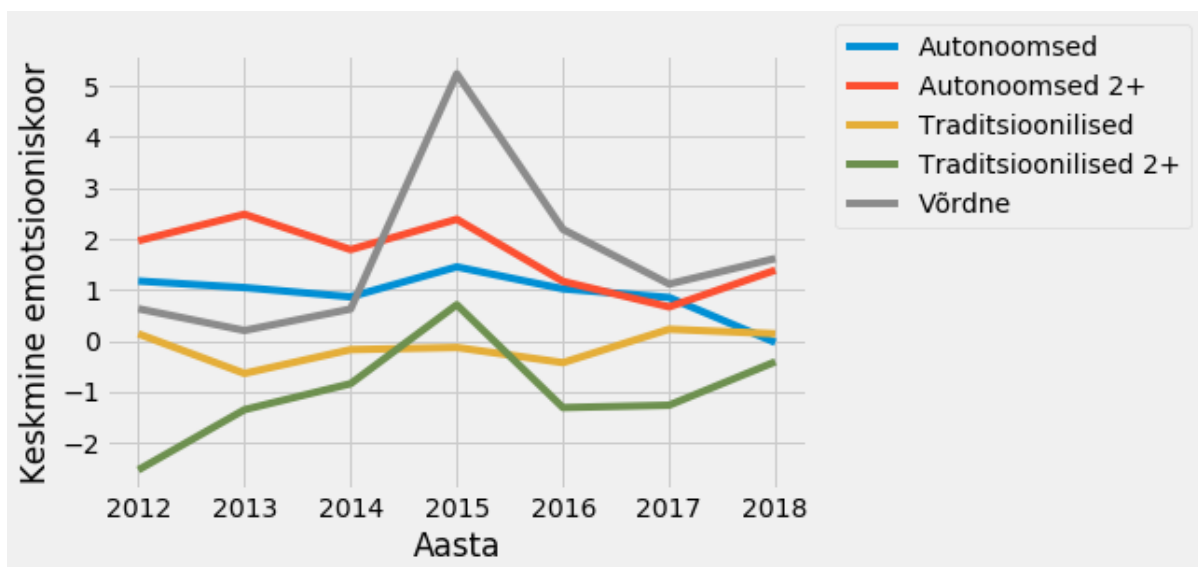
Joonis 19. Kasvatusväärtuste jagunemine tekstides väärtusjaotuse järgi

Jooniselt 19 nähtub, et väärtusjaotus on traditsioonilistel ja autonoomsetel kasvatusväärtustel samas suurusjärgus – tekste, mis sisaldavad ühte liiki väärtusi rohkem kui teisi, on pea võrdselt (autonoomseid kasvatusväärtusi ühe võrra rohkem 1981, traditsioonilisi ühe võrra rohkem 1904, autonoomseid kahe ja enama võrra rohkem 301, traditsioonilisi kahe ja enama võrra rohkem 301). Tekste, kus traditsioonilisi ja autonoomseid kasvatusväärtusi on võrdselt, leiti 123.



Joonis 20. Väärtusjaotuse keskmine emotsiooniskoor (väärtusjaotuse tekstide emotsiooniskooride summa jagatuna väärtusjaotuses olevate tekstide arvuga)

Kasvatusväärtuste keskmise emotsiooniskoori arvutamiseks jagati väärtusjaotuses olevate tekstide emotsiooniskooride summa väärtusjaotuses olevate tekstide arvuga, näiteks kui väärtusjaotuses „Vördne“ on 10 teksti, mille emotsiooniskooride summa on 20, siis on väärtusjaotuse keskmine emotsiooniskoor 2.



Joonis 21. Väärtusjaotuse keskmine emotsiooniskoor aastate lõikes

Jooniselt 18 saime indikatsiooni, et autonoomsete kasvatusväärtuste keskmine emotsiooniskoor on kõrgem kui traditsioonilistel väärtustel ja joonis 20 kinnitab meie hüpoteesi, et traditsiooniliste kasvatusväärtuste emotsiooniskoor on madalam kui autonoomsete kasvatusväärtuste emotsiooniskoor. Jaotus „Võrdne“ on suure positiivse emotsiooniskooriga ehk siis tekstides, kus on koos võrdselt nii traditsioonilised kui autonoomsed kasvatusväärtused, on neid valdavalt kajastatud positiivse emotsionaalsusega. Parema ülevaate saamiseks vaadati väärtusgruppide emotsiooniskoori ka aastate lõikes (joonis 21).

Jaotus „Võrdne“ on selgelt eristuva emotsiooniskooriga just aastal 2015, mille põhjustas asjaolu, et kuna selles jaotuses oli kõige vähem tekste, muutis üks suure emotsiooniskooriga (47) tekst arvestatavalt emotsiooniskoori. Joonisel 20 ja joonisel 21 olevate andmete põhjal saame öelda, et mida rohkem on tekstis ühe jaotuse kasvatusväärtusi võrreldes teise jaotuse omadega, seda kõrgem (autonoomsed kasvatusväärtused) või madalam (traditsioonilised kasvatusväärtused) emotsiooniskoor tekstil on.

4.6 Võimalus tekste kvalitatiivselt analüüsida

Lisaks andmete kvantitatiivsele analüüsile on sotsiaalteadlaste jaoks oluline ka tekste kvalitatiivselt analüüsida. Ainult emotsiooniskoori järgi (emotsionaalsust sisaldavate tekstiühikute väärtuste summa) tekste otsides tulevad valikusse pikemad tekstid, ainult tekstide keskmise emotsiooniskoori järgi otsides (emotsiooniskoor jagatuna sõnade arvuga) tulevad jällegi lühemad tekstid, sest mida lühem tekst, sest suurem on emotsioonisõnastikus oleva sõna (või ka tujukuju või emotsioonisõnastikus mitteoleva tegusõna) mõju keskmisele.

Paremate otsingutulemuste saavutamiseks loodi otsingutulemuste põhjaliku testimise tulemusena kaalutud keskmine emotsiooniskoor, mis arvutatakse sõltuvalt sõnade arvust tekstis järgmistel põhimõtetel:

- Kuni 10 sõna tekstis – emotsiooniskoor (emotsionaalsust sisaldavate tekstiühikute väärtuste summa) korrutatakse läbi arvuga 0,25;
- Rohkem kui 10 sõna ja vähem kui esimeses kvartiilis - emotsiooniskoor korrutatakse läbi arvuga 0,5;
- Sõnade arv vahemikus esimene kvartiil (I kv) ja III kvartiil (III kv) - emotsiooniskoor korrutatakse ühega;
- Sõnade arv suurem kui III kvartiil ja väiksem kui maksimum (kolmanda ja esimese kvartiili vahe ehk IQR, mis on korrutatud 1,5-ga + kolmas kvartiil) – emotsiooniskoor korrutatakse läbi arvuga 1,5;
- Maksimumist suurem sõnade arv - emotsiooniskoor korrutatakse läbi arvuga 2.

GitHubis on neli näidispäringut, kus kahes esimeses on emotsiooniskoori ja kaalutud keskmist emotsiooniskoori kasutades leitud viis kõige madalama emotsiooniskooriga teksti.

Kolmandas ja neljandas näidispäringus leiame üles tekstid, kus sisalduvad nii konformsed kui loomingulised väärtused. Kolmandas näidispäringus leiame kaalutud keskmist emotsiooniskoori kasutades viis kõrgema skooriga teksti ja neljandas viis madalama skooriga teksti (kood GitHub'is²⁶).

GitHubist näidispäringutega leitud tekste lugedes võib näha, et kaalutud keskmise emotsiooniskoori kasutamine tekstide sorteerimisel annab hea tulemuse, aga tulemuslik on ka tekstide sorteerimine emotsiooniskoori alusel ning sotsiaalteadlased saavad tekstide kvalitatiivseks analüüsiks mõlemat varianti kasutada.

²⁶https://github.com/Teet1970/Magistrit-/blob/master/6_mag_analyys.ipynb

5. Kokkuvõte

Käesoleva magistritöö eesmärgiks oli luua metoodika koos tarkvaralahendusega, mis võimaldaks Perekooli foorumist Koolilapse teema alt alla laadida tekstid, leida nende hulgast kasvatusväärtusi sisaldavad tekstid, arvutada tekstide emotsiooniskoorid ning analüüsida saadud andmeid. Lisaks püstitati ka kaks hüpoteesi – kasvatusväärtuste esinemissageduste muutumise kohta aastate jooksul ja kasvatusväärtustega seotud keskmiste emotsiooniskoorige osas.

Kõik magistritöö osas püstitatud eesmärgid said täidetud:

1. Magistritöö käigus loodi veebiämblik, millega laaditi alla Perekooli foorumist 158795 teksti koos metaandmetega.
2. Loodi kasvatusväärtuste sõnastik (lisa 4), mille abil eraldati 4612 kasvatusväärtusi sisaldavat teksti (2,9% tekstide koguarvust).
3. Kasvatusväärtusi sisaldavatele tekstidele arvutati emotsiooniskoorid.
4. Analüüsiti andmeid, sh loodi kaalutud keskmine emotsiooniskoor, et oleks võimalik tekste paremini kvalitatiivselt analüüsida.

Esimeses hüpoteesis eeldati, et autonoomsete kasvatusväärtuste osakaal võrreldes traditsiooniliste kasvatusväärtustega on analüüsitava ajavahemikul (2012-2018) suurenenud. Jooniselt 16 (lk. 34) nähtub, et kui 2012. aastal oli autonoomseid kasvatusväärtusi 49,3% loetud kasvatusväärtuste koguarvust, siis aastal 2015 küll korras kasvas autonoomsete kasvatusväärtuste protsent 55,7 peale, aga hiljem langes taas ja viimasel uuritud aastal oli see 46,7%, nii et esimene hüpotees ei leidnud kinnitust.

Teises hüpoteesis eeldati, et traditsioonilisi kasvatusväärtusi hindavad vanemad on rohkem mures laste käitumise pärast ning seetõttu on traditsioonilistel kasvatusväärtustel keskmised emotsiooniskoorid madalamad kui autonoomsetel kasvatusväärtustel. Joonised 18, 20 ja 21 kinnitavad selle hüpoteesi paikapidavust, kusjuures kui ühte tüüpi kasvatusväärtusi on tekstis rohkem kui teisi, siis rohkem autonoomseid kasvatusväärtusi võrreldes traditsiooniliste kasvatusväärtustega tähendab keskmise emotsiooniskoori tõusu ja rohkem traditsioonilisi kasvatusväärtusi võrreldes autonoomsete kasvatusväärtustega tähendab keskmise emotsiooniskoori langust.

Magistritöö käigus loodud metoodikat ja tarkvaralahendust on plaanis tulevikus kasutada ka teiste meediatekstide analüüsiks ning juba saadud andmete baasilt on plaan kirjutada vähemalt üks teadusartikkel. Kui Perekooli foorumi haldajalt oleks tulevikus võimalik saada anonümiseeritud kasutajate vanused tekstide kirjutamise hetkel, oleks võimalik olemasolevat analüüsi osa veelgi edasi arendada, vaadeldes kasvatusväärtuste esinemist postitajate vanuserühmade kaupa.

6. Viidatud kirjanduse loetelu

Alwin D. (2001). Parental values, beliefs, and behavior: A review and promulga for research into the new century. *Children at the Millennium: Where Have We Come From, Where Are We Going?* 6, 97–139.

Brooke J, Tofiloski M, ja Taboada M. (2009). Cross-linguistic sentiment analysis: From English to Spanish. *Proceedings of the 7th International Conference on Recent Advances in Natural Language Processing*, 50-54. Borovets, Bulgaria.

Esuli A ja Sebastiani F. (2006). SentiWordNet: A publicly available lexical resource for opinion mining. *Proceedings of 5th International Conference on Language Resources and Evaluation (LREC)*, 417-22. Genoa, Italy.

Fjellvang T. (2011). Socialization values, cultural-religious zones and modernization theory. *European Sociological Review*, 27(2), 196-211.

Kalmus V. (1999). Väärtused Eesti aabitsais käesoleval sajandil. J. Mikk (toim), *Väärtuskasvatus õppekirjanduses*, 23–37. Tartu: Tartu Ülikooli pedagoogika osakonna väljaanne nr 8.

Kalmus, V. (2019). Intervjuu 28.03.2019 Tartus.

Kerge K, Pajupuu H, Alp P, Põlda H ja Uusen A. (2014). Towards Sophisticated Writing. *Proceedings of the Tallinn University Institute of Estonian Language and Culture: Studies in Language Acquisition, Learning, and Corpora*, 16, 103–115.

Kohn, M. (1969). *Class and Conformity: A Study in Values*. Homewood, IL: The Dorsey Press.

Kwartler, T. (2017). *Text mining in practice with R*. Hoboken, NJ: John Wiley & Sons.

Mikolov T, Chen K, Corrado G ja Dean J. (2013). Efficient estimation of word representations in vector space. *Proceedings Of Workshop At ICLR*.

Orasmaa S, Petmanson T, Tkachenko A, Laur S ja Kaalep H-J. (2016). EstNLTK—NLP Toolkit for Estonian. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.

Orasmaa S ja Särg D. (2017). Kursuse „Eesti keele töötlus Pythoniga“ materjalid. TÜ ATI.

Orav H, Zupping S ja Vare, K. (2014). Leksikosemantiliste suhete hägusus Eesti Wordnetis. *Emakeele Seltsi Aastaraamat*, 60, 171–194.

Pajupuu H, Altrov R ja Pajupuu J. (2016). Identifying polarity in different text types. *Folklore. Electronic Journal of Folklore*, 64, 25–42.

Pang B ja Lillian L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1–135.

Rokeach M. (1973). *The Nature of Human Values*. New York: The Free Press.

Schwartz, S H. (2012). An Overview of the Schwartz Theory of Basic Values. *Online Readings in Psychology and Culture*, 2(1).

Schwartz S H ja Bilsky W. (1987). Toward a universal psychological structure of human values. *Journal of Personality and Social Psychology*, 53(3), 550-562.

Smith P B ja Schwartz S H. (1997). Values. *Handbook of cross-cultural psychology*, 3, 77-118. Boston: Allyn & Bacon.

Taboada M, Brooke J, Tofiloski M, Voll K ja Stede M. (2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, 37(2), 267–307.

Taboada M. (2016). Sentiment Analysis: An Overview from Linguistics. *Annual Review of Linguistics*, 2, 325–347.

Tulviste T ja Mizera L. (2010). Socialization values of mothers of Estonian preadolescents. *European Journal of Developmental Psychology*, 7(3), 295—308.

Tulviste T, Mizera L ja De Geer B. (2012). Socialization Values in Stable and Changing Societies: A Comparative Study of Estonian, Swedish, and Russian Estonian Mothers. *Journal of Cross-Cultural Psychology* 43, 480-497.

Lisad

Lisa 1. Programmeerimiseks vajaliku keskkonna seadistamine

Tarkvaralahenduse tegemiseks on kasutatud programmeerimiskeele Python versiooni 3.5, mis ei ole kõige uuem versioon, aga magistritöö raames loodud tarkvara kasutamises olulist rolli omav teek EstNLTK versioon 1.4.1 ei tööta uuemate Pythoni versioonidega.

Tarkvaralahenduse loomisel kasutati vabavaralist Anaconda Distribution tarkvarapaketti²⁷ ning kodeerimine toimus veebipõhist kodeerimiskeskkonda Jupyter Notebook'i²⁸ kasutades. Kuigi installeeritud Anaconda (vajab kettaruumi 3GB) omab graafilise liidesega keskkonda Anaconda Navigator, on otstarbekam kasutada käsureakeskkonda conda²⁹ (Windowsi kasutajatele Anaconda Prompt, macOS'i või Linux-i kasutajatele terminali aken), mis võimaldab kiiremini ja lihtsamalt soovitada vajalikke toiminguid nagu uute keskkondade loomine ja teekide (tarkvarapakettide) installeerimine.

Käsk `conda create --name minu_keskkond python=3.5.3` loob keskkonna, kus on kasutusel Python'i versioon 3.5.3, mis sobib magistritöös loodud tarkvaralahenduse kasutamiseks.

Juba loodud keskkonna aktiveerimiseks on käsk:

Windows: `activate minu_keskkond` (sulgemine `deactivate minu_keskkond`)

Linux, Mac: `source activate minu_keskkond` (sulgemine `source deactivate minu_keskkond`)

Aktiveeritud keskkonda installeeritud pakette saab vaadata käsuga `conda list`. Magistritöös loodud koodi kasutamiseks peavad olema installeeritud teegid `pandas`, `numpy`, `matplotlib`, `seaborn`, `genism`, `requests`.

Programmeerimises vajalike lisateekide installeerimiseks on vaja internetiühendust ja seda saab teha kahel viisil:

1. `conda install teegi_nimi[==versioon]` # kui teek on Anaconda repositooriumis;
2. `pip install teegi_nimi[==versioon]` # kui teek on PyPA (Python Packaging Authority) repositooriumis.

EstNLTK teegi³⁰ installeerimiseks windows-10 keskkonda tuleb sisestada käsk:

```
conda install -c estnltk -c conda-forge estnltk=1.4.1
```

²⁷ <https://www.anaconda.com/distribution/>

²⁸ <https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/>

²⁹ <https://docs.conda.io/projects/conda/en/latest/user-guide/>

³⁰ <https://estnltk.github.io/estnltk/1.4.1/tutorials/installation.html>

Ajaväljendite tuvastamiseks ning süntaksianalüsaatori ja osalausestaja kasutamiseks EstNLTK 1.4.1 teeki kasutades (antud magistritöös neid funktsioone ei kasutata) on vaja veel installeerida Java SE Runtime Environment³¹ (versioon ≥ 1.8).

³¹<https://www.java.com/en/download/>

Lisa 2. Veebiämbliku mõiste ja kasutamise viisakusreeglid

Tarkvara, mis võimaldab veebilehtedelt automaatselt infot otsida ja seda alla laadida, nimetatakse veebiämblikuteks (*web spyder*, *web crawler*). Pythoni programmeerimiskeeles on võimalik kasutada veebiämblike ehitamiseks raamistikke, näiteks Scrapy't³² või ehitada veebiämbik³³, kasutades veebilehtede allalaadimiseks teeki *requests*³⁴ ja sealt info eraldamiseks (parsimiseks) teeki *BeautifulSoup*³⁵ (installeeritud koos EstNLTK teegiga).

Veebisuhtluses kehtivad viisakusreeglid ja veebiämblikute kasutamisel tuleks neid jälgida (Orasmaa ja Särg, 2017):

1. Veebiämblik suudab teha kiirelt väga palju päringuid, koormates nii serveri üle. Viisakas on lisada päringute tegemiste vahele ajaline paus, mis ei oleks väiksem kui 1 sekund.
2. Veebilehtede omanikud koostavad tihtipeale juhendid, kus on kirjas, mida veebiämblikud tohivad teha ja mida mitte ning paiknevad need juhised failis robots.txt³⁶. Soovitav oleks kirjapandud juhiseid järgida.
3. Veebiämbliku identiteedi määramine. Iga HTTP päringu päises on info selle kohta, kes päringuid teeb. Kui päringu tegija ja eesmärk on määratlemata, tekitab see kahtlusi. Soovitav oleks määrata programmis päringu tegija identiteet ja allpool on näide, kuidas teda tehakse teeki *requests* kasutades. Selleks tuleb päises olev info uuendada päringu tegija kohta käiva infoga:

```
headers.update(  
    { 'User-Agent': 'my crawler (minu.emaili_aadress) / on  
    studying purpose', }  
)
```

4. Enne suuremahulist andmete allalaadimist oleks viisakas võtta ühendust veebilehe haldajaga ja informeerida nii soovist andmeid alla laadida kui ka anda teada andmete kasutamise eesmärkidest.

³²<https://www.digitalocean.com/community/tutorials/how-to-crawl-a-web-page-with-scrapy-and-python-3>

³³<https://www.dataquest.io/blog/web-scraping-tutorial-python/>

³⁴<http://docs.python-requests.org/en/master/>

³⁵<https://www.crummy.com/software/BeautifulSoup/>

³⁶<http://www.robotstxt.org/robotstxt.html>

Lisa 3. Küsimustik emadele, lehekülg 3 (Tulviste ja Mizera, 2010)

III III Palun kirjuta punktiirjoontele, milliseid iseenda peamisi väärtusi tahaksid oma lapsele edasi anda:

1.
2.
3.

VI Siin on loetletud mõned kasvatuspõhimõtted. Palun loe need läbi ja hinda seejärel väidete õigsust 4-palli süsteemis, kus

1 = "kindlasti mitte", **2** = "pigem mitte", **3** = "mõnevõrra" ja **4** = "kindlasti soovin"

Palun kirjuta vastused punktiirjoontele.

Soovin:

1. kasvatada oma lapsest kohusetundliku ja usaldusväärse inimese.
2. anda oma lapsele võimalikult palju tegutsemisvabadust.
3. kasvatada oma lapsest viisakas ja meeldiv suhtleja.
4. et minu laps oleks materiaalselt kindlustatud.
5. et minu laps oleks optimistlik.
6. et minu laps saaks liidrirollis olla.
7. et minu laps usuks oma võimetesse.
8. et minu lapsel oleks kujutlusvõimet ja fantaasiat.
9. õpetada oma last ennast igas olukorras usaldama.
10. et minu lapsest peetaks lugu.
11. oma lapsele võimalikult suurt otsustusõigust.
12. kasvatada oma last enesekindlaks ja iseseisvaks.
13. et minu laps pingutaks oma eesmärkide elluviimise nimel.
14.et minu laps oleks elurõõmus.
15. et minu lapsest kasvaks mõjukas inimene.
16. et minu laps austaks vanemaid inimesi.
17. õpetada oma last ise endale sihte seadma ja eesmärke valima.
18. et minu lapsest kasvaks tubli ja töökas inimene.
19. et minu laps oleks sõnakuulelik.
20. et minu laps oleks arukas ja taiplik.
21. et minu laps tunneks alati huvi maailma asjade vastu.
22. et minu laps oleks edukas.
23.et minu laps oleks õnnelik.

Lisa 4. Kasvatuseväärtuste sõnastik

Traditsioonilised kasvatuseväärtused			Autonoomsed kasvatuseväärtused	
Konformsus	Heatahtlikkus	Kohustumine	Saavutamine	Loomingulisus
kuulekas	ausus	töökas	ettevõtlikkus	loominguline
kuulekus	sõnapidaja	töökus	ettevõtlik	loovus
sõnakuulelik	sõnapidamine	püüdlik	tegu	loov
sõnakuulekus	abivalmis	püüdlikkus	ambitsioonikas	loomingulisus
sõnakuulekas	andestav	usinus	sihiteadlik	enesekindlus
sõnakuulelikkus	lojaalne	hoolas	sihikindel	enesekindel
distsipliin	isetu	vastutustunne	sihiteadlik	eneseteadvus
distsiplineeritud	altruistlik	kohusetunne	sihikindlus	eneseaustus
distsiplineeritus	aus ³⁷	kohusetundlik	järjekindel	enesesaldus
enesedistsipliin		kohusetruu	järjekindlus	eneseteadlik
austus			saavutusvajadus	uudishimulikkus
austama			visadus	uudishimulik
austav			edasipüüdlik	iseseisvus
hastikasvatatud			edasipüüdlikkus	iseseisev
korraarmastaja			algatusvõimeline	õppimistahe
korraarmastus			algatusvõime	teadmisjanuline
lugupidav			sitkus	otsustamisjulgus
respekteeriv			eesmärgikindlus	sõltumatus
korralikkus			eneseteostus	kreatiivne
viisakus			teotahteline	kreatiivsus
viisakas ³⁸				
korralik ³⁹				
sõna kuulama				
kuulama sõna				
käsku täitma				
kohustust täitma				

³⁷ Omadussõnu „viisakas“, „korralik“ ja „aus“ arvestati edaspidi kasvatuseväärtuste kontekstis ainult siis, kui neile järgnesid nimisõnad „käitumine“, „laps“, „õpilane“, „koolilaps“, „poiss“, „tüdruk“, „komme“, „kasvatus“, „kasvatamine“.

³⁸ Vaata joonealust märkust nr. 38

³⁹ Vaata joonealust märkust nr. 38

Litsents

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Teet Kalmus,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Metoodika kasvatusväärtuste automaatseks eraldamiseks *online*-meediatekstidest“, mille juhendajad on Dage Särg, MA ja prof. Veronika Kalmus, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commonsi litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Teet Kalmus

14.05.2019